Psychology and Aging
2018, Vol. 33, No. 1, 182–194

0882-7974/18/$12.00 http://dx.doi.org/10.1037/pag0000195

Modeling Age Differences in Effects of Pair Repetition and Proactive Interference Using a Single Parameter

Joseph D. W. Stephens
North Carolina A&T State University

Amy A. Overman
Elon University

In this article, we apply the REM model (Shiffrin & Steyvers, 1997) to age differences in associative memory. Using Criss and Shiffrin's (2005) associative version of REM, we show that in a task with pairs repeated across 2 study lists, older adults' reduced benefit of pair repetition can be produced by a general reduction in the diagnosticity of information stored in memory. This reduction can be modeled similarly well by reducing the overall distinctiveness of memory features, or by reducing the accuracy of memory encoding. We report a new experiment in which pairs are repeated across 3 study lists and extend the model accordingly. Finally, we extend the model to previously reported data using the same task paradigm, in which the use of a high-association strategy introduced proactive interference effects in young adults but not older adults. Reducing the diagnosticity of information in memory also reduces the proactive interference effect. Taken together, the modeling and empirical results reported here are consistent with the claim that some age differences that appear to be specific to associative information can be produced via general degradation of information stored in memory. The REM model provides a useful framework for examining age differences in memory as well as harmonizing seemingly conflicting prior modeling approaches for the associative deficit.

Keywords: associative deficit, proactive interference, age-related memory impairment, REM model

How association-specific is the associative deficit in older adult memory? A multitude of empirical findings have corroborated Naveh-Benjamin's (2000) work demonstrating that older adults are more impaired in memory for associations than in memory for single items. Nonetheless, a frequent theoretical issue within the field of memory (and cognitive psychology more generally) concerns the question of whether specific effects must be attributed to equally specific mechanisms, as opposed to the possibility that a general mechanism can account for dissociations observed in behavioral data. A recent instance of this type of debate was initiated within the cognitive aging literature by Benjamin (2010), who argued that the apparent specificity of age-related deficits in context memory could be accounted for by a general decline in memory fidelity. According to this approach, termed the DRYAD model (Density of Representations Yields Age-related Deficits), the generally sparse nature of context encoding causes contextual information to be more affected by loss of memory fidelity than item information, which is typically more densely encoded than contextual information (subject to the demands of a particular task).

In a recent exchange, Smyth and Naveh-Benjamin (2016; Naveh-Benjamin & Smyth, 2016) and Benjamin (2016) debated whether the DRYAD approach could adequately account for the associative deficit. We do not attempt to resolve this debate here, but it is nonetheless useful in light of the debate to explore how established models of associative memory might account for observed age differences. One prominent model, REM (Retrieving Effectively from Memory; Shiffrin & Steyvers, 1997), has been used to describe data from a wide variety of memory tasks in young adults, but has not been as extensively applied to the study of age-related memory change. In this article, we consider a version of REM that has been used to describe associative memory performance in young adults, and examine how the model might account for age differences in associative memory as a consequence of variation in a single, general memory encoding parameter.

The Criss and Shiffrin (2005) Paradigm and the ADH

In a study of young adult memory, Criss and Shiffrin (2005) carried out an experiment that tested associative recognition for pairs that varied in the strength of both item and associative information, as well as in their encoding context. This was achieved by using two study lists that each consisted of word-face pairs, with some of the words and faces being presented on both

Joseph D. W. Stephens, Department of Psychology, North Carolina A&T State University; Amy A. Overman, Department of Psychology, Elon University.

Joseph D. W. Stephens is supported by Air Force Research Laboratory and OSD under agreement FA8750-15-2–0116. Amy A. Overman is supported by NIH Grant R15AG052903. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NIH, Air Force Research Laboratory and OSD, or the U.S. Government. The authors thank Leslie Hart, Ben Schwartz, and Garrett Tante Burns for their helpful assistance with data collection. Some of the data reported here were presented at the 2012 Cognitive Aging Conference, Atlanta, Georgia.

Correspondence concerning this article should be addressed to Joseph D. W. Stephens, Department of Psychology, North Carolina Agricultural and Technical State University, Greensboro, NC 27411. E-mail: jdstephe@ncat.edu

lists. Among the words and faces presented on both lists, some were presented in the same pairs (a condition we will refer to as Lists 1&2 Same) and others were presented in different pairs (i.e., Lists 1&2 Different). Thus, pairs from these conditions were equated in terms of item familiarity, but differed in the familiarity of their specific pairing. In the other conditions, words and faces were presented in pairs only once, on either the first (List 1) or second (List 2) study lists. In the test phase of the experiment, intact and rearranged pairs were presented, and participants were instructed to endorse as "old" only the intact pairs that had been presented in the second study list.

With this design, the Criss and Shiffrin (2005) paradigm combined item, associative, and contextual factors into a single recognition task. Thus, for example, effects of contextual information can be tested by comparing the List 2 condition to the List 1 condition, effects of item information can be tested by comparing Lists 1&2 Different to List 2 (since items are repeated in the former condition but not the latter), and effects of associative information can be tested by comparing Lists 1&2 Same to Lists 1&2 Different (since both conditions have repeated items but only the former has repeated pairs). In other studies, separate tasks have often been used to compare memory for different aspects of studied stimuli (such as items vs. pairs, e.g., Naveh-Benjamin, 2000). Thus, the ability to examine item, associative, and contextual factors together in a single task makes the Criss and Shiffrin design useful for the investigation of age-related memory differences, which include the associative deficit as well as difficulty with source/contextual information (e.g., Bayen, Phelps, & Spaniol, 2000). Overman and Becker (2009) repeated Criss and Shiffrin's experiment with both young and older adult participants. In addition to replicating Criss and Shiffrin's basic pattern of results for young adults, Overman and Becker found that older adults' old/new pair discrimination benefited less from pair repetition than that of young adults. That is, an age-related interaction was observed such that young adults' performance was highest in the Lists 1&2 Same condition, but older adults' performance was not significantly higher in the Lists 1&2 Same condition than in the Lists 1&2 Different condition. Because the Lists 1&2 Same condition provided the greatest amount of associative information for recognition of intact pairs, the fact that this condition exhibited the greatest age difference was interpreted by Overman and Becker as consistent with the associative deficit hypothesis.

A REM Model of Associative Recognition and Aging

In addition to their experiment, Criss and Shiffrin (2005) also presented a model to describe their young adult data. The model was a version of REM in which associative features were encoded along with item and context features for each word–face pair. Although Overman and Becker (2009) extended Criss and Shiffrin's experimental paradigm to older adults, they did not attempt to use the model to account for the age differences they observed. Thus, for the current purpose of modeling age difference, we begin by applying the Criss and Shiffrin model to the Overman and Becker data set.

REM is a global matching model of memory in which to-be-remembered stimuli are encoded in memory as episodic traces, which are represented by vectors of features drawn from a geometric distribution. Encoding is assumed to be incomplete and inaccurate, so that each trace stored in memory only contains partial feature information for the stimulus it represents. At retrieval, a test stimulus is compared to each of the available episodic traces to compute a likelihood ratio, and the likelihood ratios are combined across the entire set of traces to compute the odds that the test stimulus was studied. In Criss and Shiffrin's (2005) associative version of REM, memory traces are assumed to contain features representing each of the items in a pair, features representing the current experimental context (i.e., the study list), and features that represent the unique association between the two items in the pair. Associative features are also assumed usually to have a slightly lower probability of being encoded than item or context features. At test, the associative recognition decision is based both on the odds that each individual item in a test pair was studied in the to-be-included context, as well as the odds that the particular association was studied (full details of the model are described in the Appendix).

An important aspect of the architecture of REM is that because features are drawn from a geometric distribution, some feature values are less probable than others. If a test stimulus matches a memory trace on a relatively improbable feature value, the feature match provides stronger evidence that the stimulus was studied—and is thus more diagnostic—than if the test stimulus matches a trace on a more probable feature value, or if features are missing. Because of this, the diagnosticity of the feature values plays a large role in the discrimination of old versus new items within the model.

One possibility for how aging might lead to general degradation of memory is that it reduces the diagnosticity of the information stored in memory. This idea of reduced diagnosticity is consistent with several conceptualizations of age-related memory decline. For example, Li, Naveh-Benjamin, and Lindenberger (2005) implemented a neural network model of the associative deficit based on the idea that age-related change in neuromodulation causes patterns of activation in the network to become less distinctive (realized in the model as a change to the gain parameter in network units' activation functions). In an approach based on the SAC model (Source of Activation Confusion; Reder et al., 2000), Buchler and Reder (2007) modeled age differences in memory in part by assuming that older adults' networks of semantic associations are more diffuse than those of young adults, making individual stimuli less unique. Finally, Benjamin's (2010) DRYAD model of age deficits in context memory implemented age-related changes as a loss of fidelity, that is, a reduction in the overall number and accuracy of encoded features. Within an architecture such as REM, any of these conceptualizations is functionally equivalent to a reduction in the diagnosticity of memory traces, although they may differ mechanistically in terms of whether the information encoded in memory is less distinctive to begin with, or becomes less distinctive through degradation (i.e., because loss of fidelity increases the likelihood that a diagnostic feature will either not be encoded at all, providing no information, or be encoded inaccurately and replaced by a different feature value, which is likely to be more generic). As described below, for our current implementations of the REM model, we applied both of these mechanistic approaches separately, and found that they have almost identical consequences for memory performance.

Simulations of Overman and Becker (2009) Data

We began by exactly replicating the model of Criss and Shiffrin (2005; see Appendix), and adjusting a minimal number of parameter values to provide a decent fit to the young adult data of Overman and Becker (2009). We then modeled the age difference in performance in two different ways. Each of these older adult models involved adjusting only a single parameter away from the values used for young adults. The first of these parameters was *g*, which is a parameter used to determine the shape of the geometric distribution from which features are drawn. Adjusting *g* upward places greater probability density at the low end of the feature distribution, so that a greater proportion of feature values are selected from among the least distinctive feature values. This makes feature vectors relatively more generic; one application in which higher values of *g* are used is in modeling memory for high-frequency words (e.g., Malmberg, Steyvers, Stephens, & Shiffrin, 2002; Shiffrin & Steyvers, 1997). Thus, adjustment of *g* corresponds fairly well to the reduced-distinctiveness assumptions in the models of Buchler and Reder (2007) and Li et al. (2005). The other REM parameter we used to model age differences was *c*, which represents the probability that a feature, if encoded, is encoded correctly. Reducing *c* roughly corresponds to the reduced-fidelity approach of DRYAD (Benjamin, 2010).

Figure 1 presents the data of Overman and Becker (2009) along with REM model fits obtained as described above, with 1000 simulated participants for each combination of parameter values. The model provides a good approximation of older adult performance, and does so based on either parameter adjustment. The qualitative pattern of particular interest is that the greatest decrement in performance is seen in the repeated-pair condition, Lists 1&2 Same. Figure 2 illustrates how discrimination performance in the different conditions changes across the possible range of values for each of the parameters *g* and *c*, when the other parameter is held constant. It can be seen that the fits obtained for the Overman and Becker (2009) data are representative of the overall effect of changing either parameter. That is, in both cases, as the diagnosticity of features in memory is reduced, performance in the Lists 1&2 Same condition declines at a greater rate than performance in other conditions. When viewed this way (as in Figure 2), it may seem obvious that performance in the Lists 1&2 Same condition should degrade faster than in other conditions: As performance in all conditions moves monotonically toward zero, the condition that starts out highest necessarily has to decline at the greatest rate. Indeed, part of the advantage of implementing this type of model is that it makes clear how a seemingly specific Age × Condition interaction effect can in fact come about through a general change that affects all conditions. Additionally, the fact that that the degradation can be produced by adjusting either of two parameters related to diagnosticity further supports the notion that the observed age differences in these data could be quite general.

Effects of Multiple Repetitions

The associative REM model of Criss and Shiffrin (2005) appears to do well in capturing age differences in the effects of pair repetition on associative recognition. Although Overman and Becker (2009) found virtually no benefit of pair repetition for older adults in their experiment, other studies that have used a greater number of repetitions per pair have found some benefit of pair repetition in older adults (Kilb & Naveh-Benjamin, 2011; Light, Patterson, Chung, & Healy, 2004; Van Ocker, Light, Olfman, & Rivera, 2017). Can the associative REM model similarly use reduced feature diagnosticity to describe age differences in an experiment with additional pair repetitions? For a

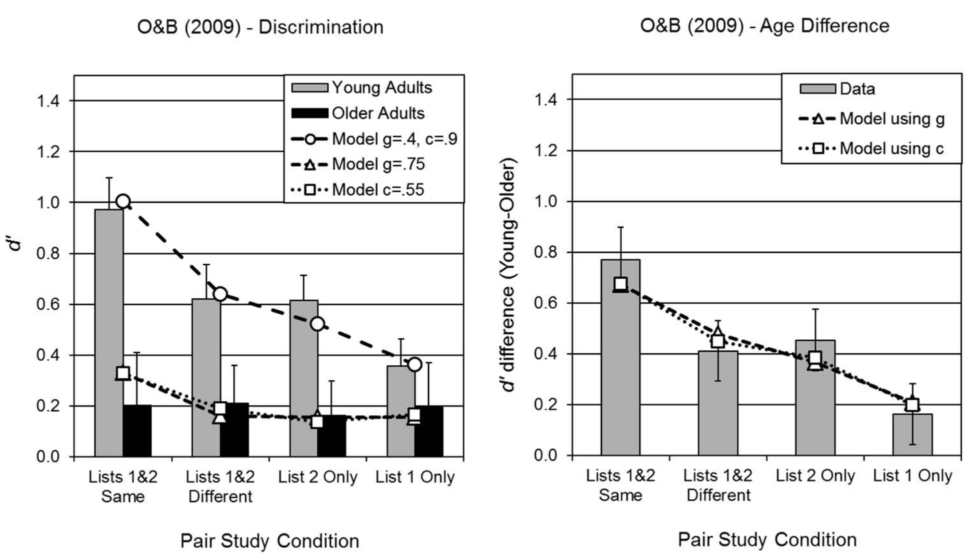

Figure 1. Overman and Becker (2009) data (bars) and model results (lines). The left panel displays mean discrimination of intact versus rearranged pairs, as indexed by d', across pair study conditions. Error bars represent standard error of the mean. The right panel displays the observed age differences (young minus older), with error bars representing standard error of the mean difference. Model fits were produced by the Criss and Shiffrin (2005) REM model with no additional assumptions. Parameters were adjusted to achieve a reasonable fit to young adult data first, then older adult data were modeled by changing only the *g* parameter or *c* parameter.

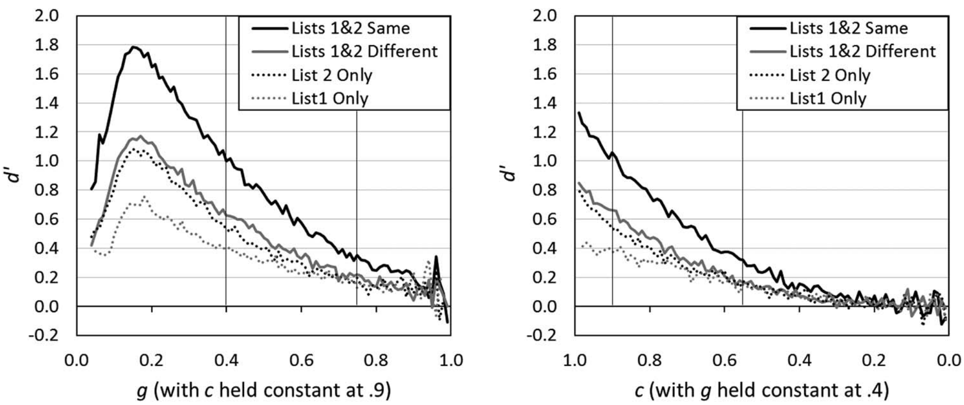

Figure 2. Model results across the range of parameter values for *g* and *c* between .01 and .99 (adjusted in increments of .01), produced by running the model with 1000 simulated participants at each parameter value. Vertical lines mark the parameter values used in the fits displayed in Figure 1. The benefit of pair repetition is seen in the difference in d' between the Lists 1&2 Same and Lists 1&2 Different conditions. Across most of the range of parameter values, reduced diagnosticity (at higher values of *g*, or lower values of *c*) diminishes performance in the Lists 1&2 Same condition at a faster rate than other conditions.

direct extension of the model, we now report data from an experiment that adapted the Criss and Shiffrin paradigm to use three study lists instead of two.

The experiment used the same basic paradigm as Criss and Shiffrin's (2005) experiment with word-face pairs, but with a third study list added. The addition of another study list also required replacement of the stimuli with a larger set (80 words and 80 faces instead of 64 of each). In exploring this variation of the paradigm, we used two alternate stimulus sets with slightly different word and face stimuli; we describe these below as Version A and Version B. The primary difference between Version A and Version B was that Version A used abstract words (like Criss and Shiffrin) whereas Version B used first names as the word stimuli.

Method

Participants

Based on similar prior experiments, participants were recruited with the goal having a sample size of at least 25 within each age group and version of the task. Thus, 57 young adults (mean age = 19.18 years, range = 18–22 years; mean education = 13.75 years, range = 12–16 years) were recruited from Elon University's introductory psychology courses and received Psychology course credit for participation. Of the 57 young adults, 32 participated in Version A (abstract words) and 25 participated in Version B (names). Fifty older adults (mean age = 74.88 years, range = 66–89; mean education = 16.77 years, range = 12–24 years), were recruited from the local community. Older adult participants were entered into a lottery to win one of 4 gift cards. Of the older adults, 25 participated in Version A of the experiment, and 25 participated in Version B. All participants were native English speakers and reported no history of major medical, neurological, or psychiatric disorders. After the explanation of procedures and prior to testing, all participants provided written informed consent to participate using consent forms approved by the Institutional Review Board of Elon University. After study participation was complete, all participants were debriefed both verbally and in writing.

Materials

The Mini Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) was administered to older adults in order to assess global cognitive function. Participants' scores on the MMSE ranged from 25 to 30 with a mean of 28.9.

For the experiment, picture stimuli consisted of 80 standardized photographs of faces. In Version A, the faces were images of 40 young females and 40 young males in front-facing position with neutral expression, selected from the FEI Face Database (Thomaz & Giraldi, 2010). In Version B, faces were selected from the Center for Vital Longevity face database (Minear & Park, 2004) and included 20 young female, 20 older female, 20 young male, and 20 older male faces with neutral expression. Word stimuli in Version A were 80 abstract words selected from the MRC psycholinguistic database (Coltheart, 1981) such that they were of similar length (M = 5.2 letters, range = 4–6 letters), imageability (M = 355.6, range = 218–437), and Thorndike-Lodge Frequency (M = 41.69, range = 12–92). Word stimuli in Version B of the experiment were 40 female and 40 male first names obtained from the Social Security database of most popular names by decade (United States Social Security Administration, n.d.). The most popular 8 names for each gender were selected from the 1990s, then each of the four prior decades (substituting from further down the list in earlier decades to avoid repeated names across decades).

Design

The study was a 5 × 2 × 2 design consisting of pair Study Condition (within-subjects) × Age Group (between-subjects) × Version (between-subjects).

The design of the study and test lists is illustrated in Table 1. There were five pair study conditions, which we refer to as Lists 12&3 Same, Lists 12&3 Different, List 3 Only, List 2 Only, and List 1 Only. The design of the study and test lists differed from that of Criss and Shiffrin (2005) only in the addition of the third study list. The List 1 Only, List 2 Only, and List 3 Only

Table 1
Pair Condition Design

Study condition	First study list	Second study list	Third study list	Test list	Test pair type	Correct response
Lists 12&3 Same	w1-f1	w1-f1	w1-f1	w1-f1	Intact	Yes
	w2-f2	w2-f2	w2-f2	w2-f3	Rearranged	No
	w3-f3	w3-f3	w3-f3	w3-f2		
Lists 12&3 Different	w4-f4	w4-f5	w4-f6	w4-f6	Intact	Yes
	w5-f5	w5-f6	w5-f4	w5-f4		
	w6-f6	w6-f4	w6-f5	w6-f5		
	w7-f7	w7-f8	w7-f9	w7-f10	Rearranged	No
	w8-f8	w8-f9	w8-f10	w8-f7		
	w9-f9	w9-f10	w9-f7	w9-f8		
	w10-f10	w10-f7	w10-f8	w10-f9		
List 3 Only			w11-f11	w11-f11	Intact	Yes
			w12-f12	w12-f13	Rearranged	No
			w13-f13	w13-f12		
List 2 Only		w14-f14		w14-f14	Intact	No
		w15-f15		w15-f16	Rearranged	No
		w16-f16		w16-f15		
List 1 Only	w17-f17			w17-f17	Intact	No
	w18-f18			w18-f19	Rearranged	No
	w19-f19			w19-f18		

Note. w1 = word1; f1 = face1, etc. In the experiment, there were equal numbers of studied pairs in all conditions (not illustrated here to conserve space). Only half of the studied List 1 Only and List 2 Only pairs were used in the test phase in order to keep the same test list length as prior studies. The correct response to intact pairs in the List 1 Only and List 2 Only conditions was "no" because participants were instructed to accept only pairs that had appeared on the third study list. Table adapted from Overman and Becker (2009).

conditions consisted of word–face pairs that appeared only once, on the applicable study list. The Lists 12&3 Different condition consisted of words and faces that appeared on all three study lists in different pair combinations on each study list. Finally, the Lists 12&3 Same condition consisted of word–face pairs that appeared on all three study lists in exactly the same combination of word and face. Each condition used 16 words and 16 faces.

create the test list. Within each condition, test items were either intact pairs that had been studied, or novel rearranged pairs created from pairs within that condition. Participants' task during the test phase was to endorse pairs that had been studied on the third study list, and to reject pairs that were new or had been studied on the first or second study lists. Thus, in the case of the Lists 12&3 Different condition, in which each word and face had been studied in a different pair on each list, intact pairs at test were identical to pairs on the third study list, and rearranged pairs at test were completely novel pairings of words and faces that had not occurred on any of the study lists. For the List 1 Only and List 2 Only conditions, the intact pairs at test were pairs that had in fact been studied, yet the correct response to these pairs was "no," because participants were instructed to say "yes" only to pairs that had been presented on the third study list. Additionally, only half of the words and faces from the List 1 Only and List 2 Only conditions were used in the test, so that the overall length of the test list remained at 64 pairs (as in prior experiments), despite the addition of the third study list.

An additional design constraint in Version B of the experiment was that names were only paired with faces of matching gender, across all lists.

Procedure

For older adult participants, the MMSE was completed prior to beginning the experiment. The experiment was conducted using E-Prime software (Psychology Software Tools, Sharpsburg, PA). Participants were seated at a computer and told that they would be seeing three lists of word-face pairs, and would later be asked to remember what they saw. Each study list was labeled with a message stating "You are beginning List 1," "You are beginning List 2," and so forth that was presented on the computer screen just before the start of each study list. Each study list contained 48 pairs of items (16 from the Lists 12&3 Same condition, 16 from the Lists 12&3 Different condition, and 16 from the relevant List 1 Only, List 2 Only, or List 3 Only condition). To maintain participants' attention to pairs during study, they were instructed to make a simple response to each pair with the keyboard. In Version A, participants indicated whether they thought the word and face went together (yes/no); in Version B, participants indicated whether they found the face-name pair pleasant or not (yes/no). Pairs were presented until participants pressed a key or for a maximum of 5000 ms. Each study trial was separated by a 500-ms interstimulus interval (ISI). In between study lists, participants completed a backward counting by threes task for one minute. Following the third study list, participants were told that they would be seeing a list containing word-face pairs, some of which they had already encountered. They were instructed to respond "yes" during the test phase only if they saw exact pairs that were from List 3. They were instructed to respond "no" to all other pairs, including intact pairs that were from List 1 or List 2. After any questions were answered, participants were presented with the test list of 64 pairs of items. Pairs were presented until the participant made a response or for a maximum of 5000 ms (failure to respond occurred on approximately 1% of all trials across participants).

Results

Old/New Pair Discrimination

For present purposes, the dependent variable of interest was old/new pair discrimination as represented by d' (Macmillan & Creelman, 2005). "Yes" responses to intact pairs (i.e., hits) and to rearranged pairs (i.e., false alarms) within each condition were used to compute d' for each participant (note that in the List 1 Only and List 2 Only conditions, the hit rate used to compute d' was also based on "yes" responses to intact List 1 Only and List 2 Only pairs, respectively, despite the fact that these responses were technically incorrect due to the list-discrimination instructions). Trials in which no response was made before the deadline were not included in the computation of hit and false alarm rates. Figure 3 displays mean d' across pair study conditions for young and older adult participants. These means were compared in a 5 (pair study condition) × 2 (age group) × 2 (experiment version) mixed-model ANOVA. There was a significant main effect of experiment version, $F(1, 103) = 4.39$, $p = .039$, $\eta_p^2 = .04$, such that overall performance was slightly better in Version A (abstract words) than Version B (names). However, there was no Condition × Version or Age Group × Version Interaction (both $F < 1$), nor a three-way interaction, $F(4, 412) = 1.331$, $p = .26$. Consequently, in Figure 3 we have presented the results collapsed across experiment version. Consistent with the findings of Overman and Becker (2009), there were significant main effects of pair study condition, $F(4, 412) = 16.21$, $p < .001$, $MSE = 12.26$, $\eta_p^2 = .14$, and age group, $F(1, 103) = 20.19$, $MSE = 20.70$, $p < .001$, $\eta_p^2 = .16$, as well as a Condition × Age group interaction, $F(4, 412) = 8.49$, $MSE = 2.122$, $p = .025$, $\eta_p^2 = .03$.

Discussion and Model

Figure 3 also displays model fits from the associative REM model as applied to the structure of the three-list experiment. As described above in the approach to modeling the Overman and Becker (2009) data, parameter adjustments were first made to provide a decent fit to young adult data, and then older adult data were modeled by changing only either the g parameter or the c parameter, with very similar results for the two parameters. For both parameters, less of an adjustment was needed away from the young adult model, suggesting that older adult performance in this experiment was less degraded than in Overman and Becker (2009).[1]

The overall results of the three-list experiment, and its corresponding model, share the pattern of particular interest from the Overman and Becker (2009) data. Specifically, young and older adult performance differs most in the condition where pairs are repeated, appearing to suggest a specific deficit in encoding associative information. It is worth noting that in the current experiment, the nature of the empirical interaction differs somewhat from the interaction found by Overman and Becker. In that experiment, older adults performed no better in the repeated-list condition than in other conditions. In the three-list experiment, older adults did perform best in the repeated list condition, just as young adults did, although the benefit of list repetition was not as great for older adults as it was for young adults. This is in line with other studies that have found pair repetition benefits for older adults in experiments with multiple repetitions (Kilb & Naveh-Benjamin, 2011; Light et al., 2004). Thus, it may have been the case in the Overman and Becker experiment that older adults' overall performance fell into a range where the benefit of pair repetition was too small to detect (indeed our model in Figure 1 predicts a small benefit). An additional consideration in the current data is that the age difference in the Lists 12&3 condition could have been compressed somewhat by older adults' low level of performance in that condition. Finally, for both age groups, the model tends to overestimate discrimination performance in the Lists 12&3 Different condition and underestimate performance in the List 3 Only condition. Notwithstanding these caveats, the model does a good job of capturing the age differences in those conditions, as well as their relative magnitude in comparison with the age difference in the Lists 12&3 Same condition, as seen in the right panel of Figure 3. For current purposes, this demonstrates that an uneven pattern of age differences in associative memory can be reproduced by a general reduction in the diagnosticity of *all* features stored in memory. As in the original model, this reduction in diagnosticity can be achieved equally well through adjustment of either the shape of the feature distribution, or the rate of encoding accuracy.

Modeling Age Differences in Proactive Interference

As discussed above, in Overman and Becker's (2009) experiment as well as the three-list experiment reported here, age differences were greater in the pair-repetition condition than in any other conditions in the task. However, this pattern of age differences has not been uniformly found across all versions of the paradigm reported to date. In particular, Overman and Stephens (2013) found a somewhat different pattern depending on what encoding strategy was employed by participants. We now turn to the data from that study to consider whether its pattern of age differences can be similarly described by the associative REM model.

Overman and Stephens (2013): Strategy Effects

The experiment reported by Overman and Stephens (2013) was identical in its basic structure to that of Criss and Shiffrin (2005) and Overman and Becker (2009), but included additional instructions intended to manipulate participants' associative encoding of word-face pairs.[2] Word stimuli in the experiment were names of occupations, and participants were assigned to one of two associative-strategy conditions. In the low-association condition, participants were instructed to visualize

[1] Perhaps not surprising, as the mean age of older adult participants in the Overman and Becker (2009) study was higher (82.2 years) than in the current experiment (74.88 years). Also, the mean education level of older adults in that study was lower (13.8 years) than in the current experiment (16.77 years), which could imply a reduced cognitive reserve (e.g., Stern, 2006).

[2] For present purposes, we examine the conditions in Overman and Stephens (2013) that had "low context salience," that is, white backgrounds during both study lists as used in Overman and Becker (2009) and in the three-list experiment reported here.

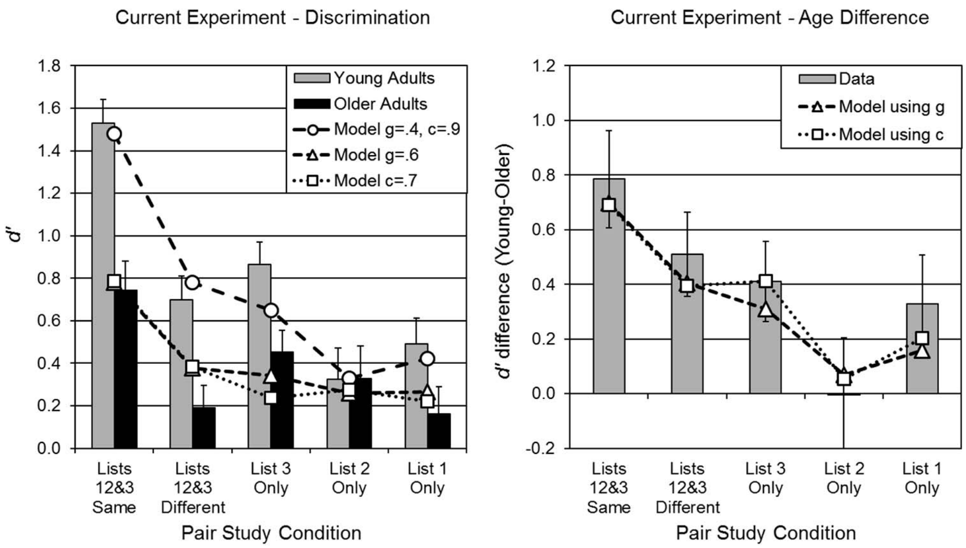

Figure 3. Results of the current experiment in which three study lists were presented (bars) with model fits (lines) from the three-list version of the Criss and Shiffrin (2005) model. The left panel presents the results in terms of mean discrimination of intact versus rearranged pairs, as indexed by d', with error bars representing standard error of the mean. The right panel displays the age difference within each condition, with error bars representing standard error of the mean difference. Model fits were again produced by adjusting parameters to fit young adult data first, then changing only the g or c parameter to fit older adult data.

each person pictured on the screen stating the name of the occupation listed next to her or his picture. In the high-association condition, participants were instructed to visualize each person *performing* the listed occupation. Predictably, the high-association strategy improved associative recognition performance in both age groups. However, the improvement was not uniform across conditions: The greatest strategy benefits were observed in young adults, in the Lists 1&2 Same and List 2 Only conditions (data are presented in Figure 4). In contrast, strategy had little effect on young adult performance in the Lists 1&2 Different condition, suggesting that there was a trade-off: Better associative encoding led to an interference effect such that intact, to-be-endorsed pairs (i.e., from the second study list) were rejected at test because of the alternate associations created during the first study list. The Lists 1&2 Same and List 2 Only pairs were immune from interference because their constituent faces and occupations had never been studied in alternate pairings.

Relative to older adults, the pattern of strategy effects in young adults created a novel age difference: In addition to lacking the pair repetition benefit of young adults, older adults in the high-association condition also lacked the proactive interference effect that reduced performance in the Lists 1&2 Different condition compared to the List 2 Only condition. Can a model describe this age difference in the interference effect using reduced feature diagnosticity, as with the pair repetition effect? The REM model we have used thus far cannot immediately answer this question, because it is not capable of producing the interference effect in the first place. The reason for this is relatively simple: The model encodes exactly the same amount of associative evidence for Lists 1&2 Different pairs as it does for List 2 Only pairs. The only difference between the two conditions is that there is an additional encoding episode for each of the individual items from the Lists 1&2 Different condition, within their alternate pairings on the first study list. However, the model includes no mechanism by which those alternate pairings on the first study list can disrupt memory for the pairings on the second study list. Thus, in order to proceed in modeling age differences in proactive interference, we must add an interference mechanism to the model.

Modeling Proactive Interference as Reencoding

How does a previously learned association disrupt memory for later associations? For our current purposes, we draw inspiration from neuroscientific studies of associative learning that have conceptualized proactive interference as arising from the activation of prior associations during the learning of new associations (e.g., De Rosa & Hasselmo, 2000; De Rosa, Desmond, Anderson, Pfefferbaum, & Sullivan, 2004). Recent findings related to proactive *facilitation* within multiple-list associative memory tasks also support the idea that encoding of pairs involves some recollection of prior pairings (Aue, Criss, & Novak, 2017; Wahlheim & Jacoby, 2013). Based on these ideas, we propose that proactive interference may result from some reencoding of prior pair information during encoding of new pairs. Within the associative REM model, this type of process can be approximated by assuming (a) that within each encoding episode, there is a probability of retrieving the memory trace for a similar, prior encoding episode, and (b) that there is some probability that features from a retrieved prior episode may be inserted into the memory trace for the current episode.

In order to implement these assumptions in the model, we included a recollection process in the encoding of memory traces

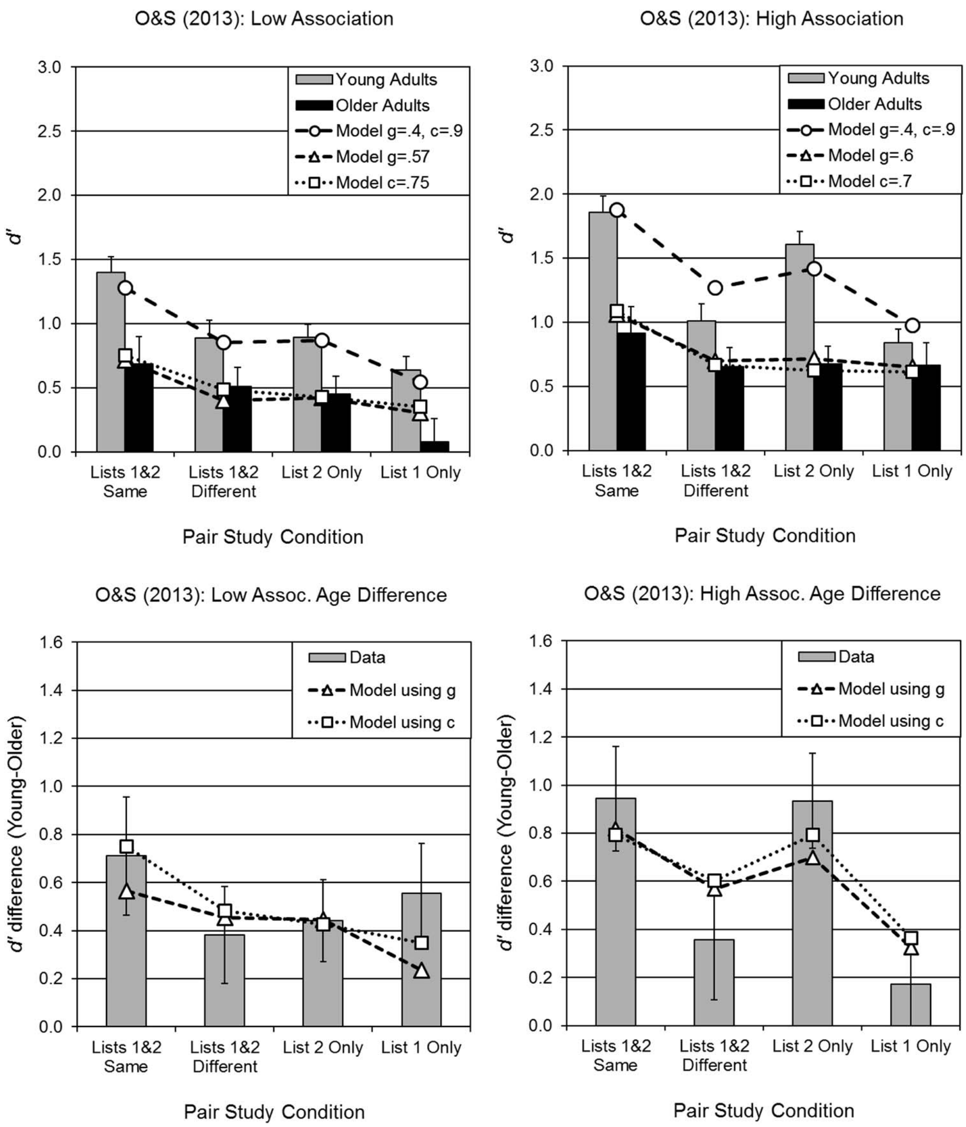

Figure 4. Data (bars) and model fits (lines) for the Overman and Stephens (2013) version of the experimental paradigm, in which participants used a low- or high-association visualization strategy to encode face-occupation pairs. Upper panels display old/new discrimination in terms of d', with error bars representing standard error of the mean. Young adults who used the high-association strategy exhibited proactive interference such that performance was worse for pairs in which the constituent items had been previously studied twice, in different pairings (Lists 1&2 Different), relative to pairs composed of items that had only been studied in a single pair (List 2 Only). Lower panels present the age difference in each condition, with error bars representing standard error of the mean difference. Model fits are results of the extended model that included a proactive interference mechanism; the modeling approach was the same as for previous models. As with the benefit of pair repetition, adjustment of either g or c alone was sufficient to reduce the effect of proactive interference in a manner consistent with the age difference.

for the second study list.[3] The recollection process is based on the sampling and recovery mechanisms of the SAM model (Raaijmakers & Shiffrin, 1980) and has been invoked (although not always fully implemented) in other versions of REM to simulate the involvement of recollection in various tasks, including single-item recognition (Malmberg, Holden, & Shiffren, 2004), associative recognition (Xu & Malmberg, 2007), and free recall (Lehman & Malmberg, 2009; Malmberg & Shiffrin, 2005).

[3] Somewhat ironically, it was not necessary to add a recollection process to the old/new decision at test. We found that doing so introduces recall-to-reject behavior that serves to reduce both hits and false alarms, having little to no effect on old/new discrimination. Thus, for the sake of parsimony, we left the retrieval process at test unchanged from that of the Criss and Shiffrin (2005) model, although a more comprehensive model of associative memory would likely include a recollection mechanism at test (cf. Xu & Malmberg, 2007).

The standard recollection process in REM uses the same likelihood values as the recognition process, computed by comparing some portion of the current stimulus (the probe) to the traces stored in memory. These likelihood values then determine the probability of each trace being sampled, and the sampled trace is then recovered with a separate probability based on how many of its features match the current stimulus (see Appendix for details).

For the current model, we let one of the items from each pair in the second study list act as a retrieval probe for memory traces from the first study list. The sampling and recovery process described above thus tends to retrieve a trace from the first study list if it contains a high match to that item. For proactive interference, we then assume that if a trace is recovered from the first study list, individual item and associative features from that trace are copied into the trace for the current pair on the second study list (note that the recovered trace is still degraded and contains an incomplete and inaccurate set of feature values). The feature values (including missing features) are copied from the recovered trace into the current trace with the same probabilities as those that determine new encoding of item and associative features from the current stimulus.

The overall effect of this mechanism is that it provides for some (relatively small) probability that a diagnostic feature of a different pair from the first study list will be copied into the trace for a pair from the second study list, thereby reducing the likelihood that the pair from the second study list will be recognized at test. Naturally, this probability is greatest in the Lists 1&2 Different condition, for which there is the best chance of recovering a prior trace that contains evidence for a different pair than the one currently being studied.

Crucially, the probability that a diagnostic interfering feature will be copied also depends on the general availability of diagnostic features in the memory traces for the first study list. One way of increasing the number of diagnostic features encoded may be through the use of an effective associative strategy. Figure 4 shows the output of the model with the proactive interference mechanism included. For the low-association strategy, we assumed a lower rate of associative feature encoding than item encoding, just as in the Criss and Shiffrin (2005) model. The overall pattern closely resembles that of the original model, with little discernible effect of interference as seen in the lack of difference between the Lists 1&2 Different and List 2 Only conditions. For the high-association model, we assumed that associative encoding occurred at the same rate as item encoding. This higher rate of associative encoding makes more features available to be reencoded into traces on the second study list, and a clear effect of proactive interference emerges (although this model somewhat undershoots the magnitude of the effect[4]).

Having incorporated interference into the Criss and Shiffrin model, we can now test whether changes in the *g* or *c* parameters alone can produce the age differences observed in the Overman and Stephens (2013) data. The lower panels of Figure 4 demonstrate that these reductions in overall feature diagnosticity do indeed cause greater degradation of performance in the Lists 1&2 Same and List 2 Only conditions than in other conditions, reproducing the qualitative pattern of age differences found in the data.

Discussion

Much of the literature on the age-related associative deficit has been based on findings that age differences are greater in memory for associative information than in memory for item information (e.g., Naveh-Benjamin, 2000). However, as Benjamin (2010, 2016) has noted, empirical interactions do not necessarily imply specific deficits. Our purpose in the present article was to examine whether some of the age interactions observed in our own current and prior work on associative memory might be accounted for by general, rather than specific, deficits in memory. The modeling results we report here indicate that some of the age differences in associative memory that we have observed, both in previously published results and in the results of a new experiment, are well described by a model in which older adult memory generally operates with less information than young adult memory overall. By "less information" we mean, quite literally, that our models of older adult performance simply used less informative (i.e., less diagnostic) features to represent stimuli than our models of young adult performance. Interestingly, more than one parameter in the model can achieve the effect of making features less diagnostic. Specifically, we achieved nearly identical results by adjusting the *g* parameter, which determines the geometric probability distribution used to generate feature values, and the *c* parameter, which determines the accuracy rate for features that are encoded in memory traces. Hence, there is more than one way to be less diagnostic.

Because of this (i.e., "more than one way to be less diagnostic"), REM yields an important insight beyond what has been provided by other models of the associative deficit hypothesis. That is, by framing the deficit in terms of feature diagnosticity, REM provides an alternate conceptualization that harmonizes seemingly conflicting approaches used in other models. For example, in their response to Benjamin's (2016) application of DRYAD to the associative deficit, Naveh-Benjamin and Smyth (2016) argued that a better approach was to be found in the neural network model of Li et al. (2005), in which impaired neuromodulation caused less distinctive representations. As noted earlier, DRYAD's concept of reduced fidelity maps on to our adjustment of REM's c parameter, whereas Li et al.'s reduction in distinctiveness is analogous to our adjustment of REM's g parameter. As such, the versions of the REM model that we describe here provide a framework for understanding the contrast between the two approaches, while also suggesting that the difference between such approaches may not be especially meaningful, insofar as they may be functionally equivalent when it comes to the diagnosticity of information in memory.

Although we have shown that the age differences examined here can be largely explained with general deficits, we do not intend to argue against the existence of specific associative deficits altogether. It may be noted that, while the model fits presented here

[4] It is possible to strengthen the proactive interference effect within the model by making additional assumptions about how features from prior pairs are copied into current pairs, for example by copying associative features with a higher probability than other features. However, we sought to add as few additional assumptions to the Criss and Shiffrin (2005) model as possible. We also note that the model of proactive interference presented here is not intended to be definitive. Our main interest is in demonstrating that reducing feature diagnosticity in the model reduces proactive interference to a similar extent as the age difference observed in the empirical data.

generally reproduce the qualitative pattern of age differences across conditions, the model does have a tendency to underestimate age differences in the repeated-list conditions relative to some of the other conditions. Thus it is likely that better fits could be obtained by including some specific associative deficits in the model, such as a reduction in associative feature encoding. We intentionally avoided adding any special assumptions to the current models as a test of how well general mechanisms alone could capture the basic patterns in the data. Additionally, the data and versions of the REM model presented here are limited to a specific experimental paradigm, and we have made no attempt to model manipulations such as divided attention, which have been used to test predictions based on general decline hypotheses (e.g., Smyth & Naveh-Benjamin, 2016). Formal process models like REM and DRYAD may be particularly useful as tools for further exploration of how various mechanistic assumptions regarding age differences may (or may not) be able to reproduce empirical patterns. Moving forward, additional useful insights into the nature of age differences may also be obtained through the application of a descriptive Bayesian approach to modeling and model selection (e.g., Tauber, Navarro, Perfors, & Steyvers, 2017).

The models presented here are also quite simplistic with regard to what they assume about retrieval processes that occur during study. Indeed, the models we present for the Overman and Becker (2009) data, and for the data from the current experiment, do not include any retrieval mechanism during study (which could account for some of the model's divergence from human performance in both age groups for the current experiment, given that study-phase retrieval might be expected to play a greater role with multiple study lists). Likewise, the study-phase retrieval mechanism that we added to the model for the Overman and Stephens (2013) data was minimal in terms of its assumptions about how prior study episodes are retrieved and reencoded. All of these choices were intentional, to avoid adding any unnecessary assumptions to the model. Nonetheless, it should be acknowledged that a complete understanding of age differences in performance on multiple-list memory tasks would include additional mechanisms for study-phase retrieval, including metacognitive mechanisms. For example, Wahlheim (2014) found that older adults are less likely than young adults to detect when pairings of stimuli change across study lists. The detection and subsequent recollection of such changes appears to support proactive facilitation in cued recall tasks, whereas the failure to do so results in proactive interference (Wahlheim & Jacoby, 2013). Thus a more detailed model might include mechanisms whereby prior stimuli that are strongly recalled during a second study list enhance memory traces in such a way that subsequent memory for the temporal order of episodes is more accurate (see also Aue et al., 2017). An interesting question for future research is whether such a model could predict the observed age differences in interference for cued recall using a reduction in feature diagnosticity as we have for the associative recognition data considered here.

In summary, the data and models presented here provide a concrete example of how empirical findings that imply a specific associative deficit can potentially be attributed to general mechanisms. These results should encourage further theoretical work that evaluates whether associative deficits found in other paradigms can be modeled with reductions in the diagnosticity of memory information. Finally, the current findings also demonstrate the usefulness of the REM model as a framework for understanding age-related memory decline. The model is well-developed and has been applied to a wide variety of memory paradigms in young adults over the last 20 years, so the opportunities are ripe for further research that examines age differences within the REM framework.

References

Aue, W. R., Criss, A. H., & Novak, M. D. (2017). Evaluating mechanisms of proactive facilitation in cued recall. *Journal of Memory and Language, 94,* 103–118. http://dx.doi.org/10.1016/j.jml.2016.10.004

Bayen, U. J., Phelps, M. P., & Spaniol, J. (2000). Age-related differences in the use of contextual information in recognition memory: A global matching approach. *The Journals of Gerontology: Series B, 55,* P131–P141. http://dx.doi.org/10.1093/geronb/55.3.P131

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review, 117,* 1055–1079. http://dx.doi.org/10.1037/a0020810

Benjamin, A. S. (2016). Aging and associative recognition: A view from the DRYAD model of age-related memory deficits. *Psychology and Aging, 31,* 14–20. http://dx.doi.org/10.1037/pag0000065

Buchler, N. E. G., & Reder, L. M. (2007). Modeling age-related memory deficits: A two-parameter solution. *Psychology and Aging, 22,* 104–121. http://dx.doi.org/10.1037/0882-7974.22.1.104

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 33,* 497–505. http://dx.doi.org/10.1080/14640748108400805

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1199–1212. http://dx.doi.org/10.1037/0278-7393.31.6.1199

De Rosa, E., Desmond, J. E., Anderson, A. K., Pfefferbaum, A., & Sullivan, E. V. (2004). The human basal forebrain integrates the old and the new. *Neuron, 41,* 825–837. http://dx.doi.org/10.1016/S0896-6273(04)00080-7

De Rosa, E., & Hasselmo, M. E. (2000). Muscarinic cholinergic neuromodulation reduces proactive interference between stored odor memories during associative learning in rats. *Behavioral Neuroscience, 114,* 32–41. http://dx.doi.org/10.1037/0735-7044.114.1.32

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

Kilb, A., & Naveh-Benjamin, M. (2011). The effects of pure pair repetition on younger and older adults' associative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 706–719. http://dx.doi.org/10.1037/a0022525

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 970–988. http://dx.doi.org/10.1037/a0015728

Li, S. C., Naveh-Benjamin, M., & Lindenberger, U. (2005). Aging neuromodulation impairs associative binding: A neurocomputational account. *Psychological Science, 16,* 445–450.

Light, L. L., Patterson, M. M., Chung, C., & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition, 32,* 1182–1193. http://dx.doi.org/10.3758/BF03196891

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Malmberg, K. J., Holden, J. E., & Shiffren, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old–new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 319–331. http://dx.doi.org/10.1037/0278-7393.30.2.319

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 322–336. http://dx.doi.org/10.1037/0278-7393.31.2.322

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition, 30,* 607–613. http://dx.doi.org/10.3758/BF03194962

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36,* 630–633. http://dx.doi.org/10.3758/BF03206543

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1170–1187. http://dx.doi.org/10.1037/0278-7393.26.5.1170

Naveh-Benjamin, M., & Smyth, A. C. (2016). DRYAD and ADH: Further comments on explaining age-related differences in memory. *Psychology and Aging, 31,* 21–24. http://dx.doi.org/10.1037/pag0000066

Overman, A. A., & Becker, J. T. (2009). The associative deficit in older adult memory: Recognition of pairs is not improved by repetition. *Psychology and Aging, 24,* 501–506. http://dx.doi.org/10.1037/a0015086

Overman, A. A., & Stephens, J. D. W. (2013). Synergistic effects of encoding strategy and context salience on associative memory in older adults. *Psychology and Aging, 28,* 654–665. http://dx.doi.org/10.1037/a0031441

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. *Psychology of Learning and Motivation, 14,* 207–262. http://dx.doi.org/10.1016/S0079-7421(08)60162-0

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 294–320. http://dx.doi.org/10.1037/0278-7393.26.2.294

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166. http://dx.doi.org/10.3758/BF03209391

Smyth, A. C., & Naveh-Benjamin, M. (2016). Can DRYAD explain age-related associative memory deficits? *Psychology and Aging, 31,* 1–13. http://dx.doi.org/10.1037/a0039071

Stern, Y. (2006). Cognitive reserve and Alzheimer disease. *Alzheimer Disease and Associated Disorders, 20,* 112–117. http://dx.doi.org/10.1097/01.wad.0000213815.20177.19

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review, 124,* 410–441. http://dx.doi.org/10.1037/rev0000052

Thomaz, C. E., & Giraldi, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing, 28,* 902–913. http://dx.doi.org/10.1016/j.imavis.2009.11.005

United States Social Security Administration. (n.d.). *Popular baby names by decade*. Retrieved from https://www.ssa.gov/oact/babynames/decades/

Van Ocker, J. C., Light, L. L., Olfman, D., & Rivera, J. (2017). Effects of repetition on age differences in associative recognition. *Memory, 25,* 350–359. http://dx.doi.org/10.1080/09658211.2016.1177089

Wahlheim, C. N. (2014). Proactive effects of memory in young and older adults: The role of change recollection. *Memory & Cognition, 42,* 950–964. http://dx.doi.org/10.3758/s13421-014-0411-4

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41,* 1–15. http://dx.doi.org/10.3758/s13421-012-0246-9

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition, 35,* 526–544. http://dx.doi.org/10.3758/BF03193292

Appendix

The Criss and Shiffrin (2005) Model of Associative Recognition

As described in the text, the REM model (Shiffrin & Steyvers, 1997) represents stimuli as vectors of features generated from a geometric distribution, such that the probability of any given feature V having value j is provided by the formula

$$P(V = j) = (1 - g)^{j-1}g. \tag{1}$$

The Criss and Shiffrin (2005) version of the model assumes that each word–face pair is represented by 45 total features: 15 for each item, 15 for context, and 15 associative features that represent the unique word–face combination. For all our models of young adult data, we generated features with Criss and Shiffrin's value of $g = .40$. The subparts of each stimulus were generated separately and combined according to the list structure used in the experiment, such that some picture features, some word features, and some associative features were repeated across study lists depending on the condition. Context features were assumed to remain constant within a study list, and to overlap between study lists due to their overall similarity; the overlap was achieved by copying some features of the List 1 context into the List 2 context. For the Overman and Becker (2009) data, we set the probability of a context feature being copied as $p_{ctx} = .70$ (same as Criss & Shiffrin).

(*Appendix continues*)

REM assumes that storage of memory traces is incomplete and inaccurate, such that each feature is stored with probability u, and if not stored, is replaced with 0. Criss and Shiffrin assumed that associative features require more effort to generate and store than other features, and thus specified a separate parameter, $u_{associative}$, for the probability of encoding an associative feature. For our model of the Overman and Becker data, we used $u = .35$ and $u_{associative} = .15$ (Criss & Shiffrin used $u = .32$ and $u_{associative} = .20$). If a feature is encoded, there is still a probability that it will be encoded incorrectly; for all of our models of young adult performance we set this probability as $c = .90$ (same as Criss & Shiffrin). Incorrectly encoded features are replaced by randomly selected feature values from the distribution.

The feature vectors generated for the study list were also used to construct a test list according to the structure of the experiment. The context features for each item were assumed to be the List 2 context features, since the task instruction was to only recognize pairs presented on the second study list. Generally speaking, REM evaluates the match between test stimuli and memory traces by computing a likelihood value λ for each memory trace according to the following equation:

$$\lambda_i = (1 - c)^{n_{iq}} \prod_j \left[\frac{c + (1 - c)g(1 - g)^{j-1}}{g(1 - g)^{j-1}} \right]^{n_{jim}}, \tag{2}$$

where n_{iq} is the number of nonzero mismatching features, and n_{jim} is the number of matching features with value j, for stimulus i. Criss and Shiffrin's associative version of the model assumes that for each test stimulus, a separate set of likelihood values λ_{iI} are computed for each group of item features in the test stimulus, as well as likelihood values λ_{iC} for the context features, and λ_{iA} for the associative features. The item and context likelihood values are combined to produce a value for the odds that the test item was studied in the relevant context:

$$\Phi_{item} = \frac{1}{N} \sum_i [\alpha \lambda_{iI}^{-1} + (1 - \alpha)\lambda_{iC}^{-1}]^{-1}, \tag{3}$$

where N is the number of memory traces over which the likelihood values were computed, and α is a parameter that weights the contribution of item and context likelihoods to the odds (set at $\alpha = .50$ in Criss & Shiffrin and all models in the current study). The associative likelihood values are used to produce a separate odds:

$$\Phi_{associative} = \frac{1}{N} \sum_i \lambda_{iA}. \tag{4}$$

The old/new decision for the test stimulus is then based on the comparison of each odds to a criterion. The model responds "old" for the entire stimulus if the item odds Φ_{item} for both the word and picture exceed the item criterion, and the associative odds $\Phi_{\text{associative}}$ exceeds the associative criterion. For our model of the Overman and Becker young adult data, we used an item criterion of 1.0 and an associative criterion of 0.9 (Criss and Shiffrin used 1.5 and 0.9, respectively).

For our models of older adult data from Overman and Becker (2009), we first found an approximate fit to the proportion old data (i.e., hits and false alarms) for young adults by adjusting parameters to the values mentioned above. We then fixed all parameter values except for either g or c, and adjusted each of those parameters separately until we found an approximate fit to the d' data for older adults. By modeling the age difference in terms of d' (instead of hit and false alarm rates) we kept our focus on modeling the age difference in old/new discrimination and avoided the need to also adjust the model's decision criteria. Model results presented in Figures 1 and 2 represent averages of $n = 1000$ simulated participants for each parameter setting, with all stimulus features, etc. generated independently for each simulated participant.

Adding a Third Study List

The model for the present experiment with three study lists had all of the same basic assumptions as the Criss and Shiffrin (2005) model for the experiment with two study lists. The number of stored traces and the construction of the study and test lists were modified according to the design of the experiment. Context features for the third study list were copied out of the context features for the second study list, after the copying was performed from the context features of the first study list into the second study list, and in the same manner, with $p_{ctx} = .70$. Test stimuli contained context features for List 3, since the task was to endorse pairs that were studied on the third study list. All other assumptions and parameters for the young adult data were identical to the model used for the Overman and Becker data, except for $u_{associative} = .20$. The modeling approach was the same: Adjustments were made to fit the young adult hit and false alarm data first, then all parameters were held constant except for g and c, which were adjusted separately to produce the two fits for older adult d' data.

(*Appendix continues*)

Adding Proactive Interference

As described in the text, we assumed that proactive interference resulted from a retrieval of List 1 traces during the study phase for List 2, which then caused some of the features of those List 1 traces to be copied into the List 2 traces. We implemented the retrieval of List 1 traces according to the search and recovery process described in Malmberg and Shiffrin (2005). For each stimulus presented in List 2, we assumed that one of the items could act as a retrieval cue for a prior trace. For simplicity, we let this be the portion of the stimulus designated as containing the picture features, but the word and picture portions of the feature vectors in this model are interchangeable. The picture features of each List 2 stimulus were compared to the stored (degraded) picture features of every List 1 memory trace to compute likelihood values according to Equation 2 above. Based on these likelihood values, the probability of sampling a given List 1 trace image, I_i, based on the picture retrieval cue, Q, was:

$$P(I_i|Q) = \frac{\lambda_i^\gamma}{\sum \lambda_i^\gamma}, \tag{5}$$

where λ_i is the likelihood ratio from Equation 2 and γ is a scaling factor; we used $\gamma = .20$ (same as Malmberg & Shiffrin). Using these probabilities, the model then sampled a trace from List 1; however, the sampled trace was not necessarily successfully recovered. The probability of successful recovery of the trace, P(R), was determined by the following equation:

$$P(R) = \rho_r^\tau, \tag{6}$$

where ρ_r represents the proportion of correctly stored features in the memory trace as a whole, and τ is another scaling factor. We used $\tau = .50$ (again, same as Malmberg & Shiffrin). If a List 1 trace was successfully recovered, we then assumed that the feature values of that trace (in its degraded form, i.e., including incorrect values and zeros) were copied into the memory trace for the List 2 stimulus with the same probabilities, u and $u_{associative}$, that were used to determine the probability of encoding stimulus features. Once the List 2 traces were modified according to this proactive interference process, all of the assumptions and retrieval mechanisms were applied as in the Criss and Shiffrin (2005) model we used for the Overman and Becker (2009) data.

The modeling procedure was the same as for the previous models. We adjusted parameters to achieve a decent fit to the hit and false alarm data for young adults in each of the associative strategy conditions. For the low-association condition, we used $u = .35$, $u_{associative} = .24$, and $p_{ctx} = .60$, with an item criterion of 1.25 and associative criterion of 0.75. For the high-association condition, we used $u = .35$, $u_{associative} = .35$, and $p_{ctx} = .60$, with an item criterion of 0.85 and associative criterion of 0.65 (the relatively low values of the criteria compensated for the overall reduction in familiarity caused by the interference mechanism). Once an acceptable set of parameters was found for the young adult data, all parameters were held constant except for g or c, which were adjusted separately to produce the fits for older adult d' data.

Received April 30, 2017
Revision received August 4, 2017
Accepted August 4, 2017 ■