Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 6, 1199–1212

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.6.1199

List Discrimination in Associative Recognition and Implications for Representation

Amy H. Criss
Carnegie Mellon University

Richard M. Shiffrin
Indiana University

Four experiments tested the predictions made by the model outlined in A. H. Criss and R. M. Shiffrin (2004b). Participants studied 2 successive lists of pairs followed by a recognition memory test for the most recent list. Some items and some pairs were repeated across the 2 lists. Critically, a given item could be repeated in the same or different type of pair. For associative recognition, performance was only affected by repetitions in the same pair type. However, in single-item recognition confusions occurred for both types of repetitions. The results are as predicted and confirm the assumption that different associative representations were stored even when the same token repeated in different pair types, whereas similar item representations were used regardless of pair type.

Keywords: list discrimination, associative recognition, memory models, recognition memory

A frequently pursued question in many domains within psychology is whether a set of features is more than a simple sum of their parts (e.g., Asch, 1964, 1969). Within the domain of human memory, this question has taken the following form: Is the association between two items stored as the simple co-occurrence of the two items or as an emergent set of features? Each of these assumptions has been adopted in extant competing models. For example, models such as REM (Shiffrin & Steyvers, 1997) and MINERVA (Hintzman, 1988) have adopted the co-occurrence assumption and represent an association as a combination of the two vectors representing each of the two singletons (a concatenation into a double long vector in REM and a summation of the two item vectors into a single vector in MINERVA). Models including TODAM (Murdock, 1982, 1997) and CHARM (Metcalfe-Eich, 1985) assume an emergent representation and model it as a third vector that contains features independent of either vector representing the singletons.

Early empirical work addressing this issue focused on paired-associate learning. Many studies were developed to uncover the conditions where the learning of the pair AC is affected by the prior learning of pairs sharing the single items A or C (e.g., AB or DC, in which the first letter represents the word given as a cue and the second letter represents the response to be generated by the participant; e.g., Greeno, James, & DaPolito, 1971; Martin, 1968; Postman, 1976). The hypothesis that pairs are stored as emergent configurations was supported when learning AC did not change (relative to baseline) following study of pairs such as AB or DC. A competing hypothesis holds that an association is simply a link or connection between two existing items in memory; in this case, the level of interference is determined by the number and strength of these links. Different sets of data favored each hypothesis. For example, some studies found positive transfer occurs when the cues are related to one another. That is, performance for AC is better following learning of BC when A and B are related (e.g., Greeno, James, DaPolito, & Polson, 1978). On the other hand, negative transfer occurs when the cue is repeated with a new response unrelated to the previous response (e.g., Greeno et al., 1971; Melton & Martin, 1972). That is, performance for AC is worse following study of AD when D and C are unrelated. This is often attributed to *persistence in encoding* or the idea that once an item is encoded in a particular way, it tends to be encoded in a similar manner in future study episodes. However, the empirical support for each of these was marginal. Studies have found the opposite of each (e.g., Greeno et al., 1978; McGeoch, 1942), and other studies found a decrement in performance when any member of the pair is repeated (Rock & Ceraso, 1964). This lack of a clearly interpretable picture surely contributed to decades of neglect of the issue in question, especially in the domain of paired-associate learning.

More recently, these issues have been addressed using the associative recognition (AR) task. In AR, participants study pairs (AB, CD, EF) and are tested with intact (AB) pairs and rearranged (CF) pairs. In a typical design, the familiarity of any individual item is irrelevant because both test items had been studied on the preceding list. To be successful in this task, participants must be able to judge whether the two items occurred together. Several studies have tried to distinguish the co-occurrence assumption and the emergent features assumption using AR, with the results typically favoring an emergent features approach (e.g., Clark, Hori, & Callan, 1993; Criss & Shiffrin, 2004b; Hockley & Cristi, 1996a,

Amy H. Criss, Center for the Neural Basis of Cognition & Department of Psychology, Carnegie Mellon University; Richard M. Shiffrin, Departments of Psychology & Cognitive Science, Indiana University.

This article was part of Amy H. Criss's dissertation. This research was supported by a Dissertation Year Fellowship from Indiana University to Amy H. Criss and National Institute of Mental Health Grant 12717 to Richard M. Shiffrin. We thank members of the dissertation committee, William K. Estes, Robert M. Nosofsky, and Joseph E. Steinmetz for many helpful discussions and comments.

Correspondence concerning this article should be addressed to Amy H. Criss, Center for the Neural Basis of Cognition, Carnegie Mellon University, 115 Mellon Institute, Pittsburgh, PA 15213. E-mail: acriss@andrew.cmu.edu

1996b; Kahana, 2002; Murnane & Shiffrin, 1991, included a sentence version of AR; see Clark & Gronlund, 1996, for a review of such studies). However, in several studies, issues of representation became entangled with issues of the processes underlying AR (i.e., whether AR requires an additional search-based recall or recollection process rather than a single familiarity process).

Criss and Shiffrin (2004b) obtained evidence pointing to differences in representation as a basis for the patterns of data, regardless of the nature of the retrieval process used to carry out AR. We mixed various classes of pairs and found that AR performance was determined by the number of pairs within one class, but not by the number of pairs in other classes. Specifically, we found that word–word pairs (WW), word–face pairs (WF), and face–face pairs (FF) did not interfere with one another, even though we found interference within each class. For example, performance for WF pairs was determined solely by the number of studied WF pairs; adding WW or FF pairs to the study list had no influence on WF performance, but adding WF pairs lowered performance. The same result held for all three types of pairs, despite some pairs sharing a common type of single item (e.g., both WF and WW pairs contain words and both WF and FF pairs contain faces). We also tested single item recognition and obtained a different result: Differences in the number of each pair type had no effect. These results are not consistent with any extant quantitative memory models because such models assume some overlap of representation (albeit for different reasons in different models) and therefore predict between-class interference. In co-occurrence models, such as REM and MINERVA, pairs are composed of the same features as the singles from which they are composed and thus both must show the same pattern of interference. In models assuming emergent pair features, such as TODAM and CHARM, pairs and singles contain different information but nonetheless are combined into a single composite memory vector, and thus all studied items contribute to the memory decision.[1] To account for the pattern of data, we suggested modifications that could be implemented in any extant model. Specifically, we suggested the following: (a) that single items are stored such that they contribute to the decision about other single items, regardless of the type of pair in which the items were studied; (b) that pair features are emergent in that they do not contain the same information as single items; and (c) that pairs of different types are stored with dissimilar representations.

In this study, we continued to explore the viability of co-occurrence and emergent assumptions by gathering evidence on the relationship between the stored associative features and the features of the items from which the association was created. To do so, we used a three-phase list discrimination design (e.g., Criss & Shiffrin, 2004a; Jacoby, 1991; Maddox & Estes, 1997) in which two lists of pairs are studied. Some singles and some pairs were repeated between the study lists, but the recognition memory test required an "old" response only if the exact test probe had been on the most recent list. Across the two lists, some participants studied repetitions of items in the same type of pair and others studied item repetitions occurring in different types of pairs. Thus, we were able to measure the contribution of item repetitions that occurred in the same or different type of studied pair. In addition to gathering converging evidence for the Criss and Shiffrin (2004b) findings and testing the proposed model, this set of experiments also provided one of the first sets of data on list discrimination in AR. Our design was based in part on prior findings showing that participants have difficulty rejecting single items that were presented on study lists other than the one being tested (e.g., Criss & Shiffrin, 2004a; Hintzman, Caulton, & Levitin, 1998). However, we know of no study that tests AR in a similar paradigm. Thus, we tested both AR and single-item recognition to compare the patterns of data between the two tasks.

Experiments 1 and 2

In our previous AR studies showing a list length effect restricted to pairs of the same type, we did not measure direct interference because individual items did not repeat (Criss & Shiffrin, 2004b). Instead, we measured interference by within-class and between-class list length effects. We found interference due to adding other pairs of the same type to the study list but no interference from other pairs of a different type and consequently no interference from other items that were presented in those pairs. For example, WF and WW pairs both contain words. The lack of cross-talk between these pairs implies that words from WW pairs do not contribute to the memory decision for a WF pair and vice versa. In the current experiments, we used a different paradigm that allows repetitions of particular items (token repetition) and thus allows stronger conclusions about the presence or absence of cross-talk between different types of pairs. The critical data come from repeating an identical item in the same or different type of pair. On the basis of the representational assumptions outlined in Criss and Shiffrin (2004b), we expected to find an interaction such that performance is affected when item repetitions occur in the same type of pair but no change in performance when an identical item is studied in a different type of pair. In the following experiments, participants studied two lists with some items and some pairs repeating in both lists. For one group of participants (Experiment 1), all item repetitions occurred in the same type of pair, and for the other group (Experiment 2), all item repetitions occurred in a different type of pair. During the surprise AR test that followed, participants were asked to accept pairs studied on the most recent list and reject all others.

Experiment 1

Method

Participants. Eighty-one people from the Indiana University community participated in the experiment in exchange for partial course credit or $7.00 per hour.

Materials. Black and white photographs of faces were selected primarily from college yearbooks and from the Olivetti Research Database of Faces (American Telephone & Telegraph, 1994). Each of the 210 faces was standardized so that the head orientation, level of the eyes, and position of the chin were identical and there was very little (if any)

[1] Note that the most recent version of TODAM (Murdock, 1997; Murdock & Kahana, 1993) assumes that context is not used for an AR decision and that the memory vector contains stored traces at the beginning of the experiment. In combination, these assumptions produce no forgetting for pairs. Earlier versions of TODAM (Murdock, 1982) did produce forgetting of pairs because of interference from other study trials. Thus, TODAM can either predict no interference for pairs using the most recent set of assumptions or it can predict interference, but it cannot predict both patterns simultaneously.

background. The 476 words varied in environmental frequency (M = 18.49, SD = 24.32 ; range = 1–245; Kucera & Francis, 1967) and were ranked low on imageability (M = 341.69, SD = 43.12; range = 129–400; Coltheart, 1981). The set of words did not include any words that might describe a face, a person, or a characteristic of either.

Procedure. Participants received two study lists separated by an unfilled break of approximately 120 s. The first study list contained 52 pairs of items and the second contained 60 pairs. On each trial of each list, participants had 3 s during which they performed an incidental task that involved rating each pair on the following question: "Do these two items go together?" Each study trial was separated by a 500-ms interstimulus interval. Following the final study list, participants were engaged in a 45-s math task before beginning an unexpected memory test. Prior to this 72 trial test list, participants were given examples of possible types of targets and foils and instructed to say "old" only to intact pairs from List 2 and to say "new" to all other pairs. Note that the above details are identical for Experiments 1 and 2, including the instructions (i.e., the example targets and foils provided to the participants were identical even if only a subset was actually tested for that participant).

Design. Both List 1 and List 2 contained all WF pairs, thus we denote this the *same* condition because items are repeated in pairs of the same type across lists. List 2 was composed of an equal number of pairs from each of the following conditions: studied only on the second list (List 2 condition), studied in exactly the same pair on Lists 1 and 2 (Lists 1 and 2 exact condition), and items studied on Lists 1 and 2 but in different pairs (Lists 1 and 2 recombined condition).[2] The test list was composed of 12 intact pairs (targets) and 8 rearranged pairs (foils) from each of these three conditions, with the foils comprised of two items studied on the second list in the respective condition but in different pairs from that condition. Twelve additional foils were constructed by testing pairs from List 1. Six of these were an exact match to a pair studied during List 1 but were foils because they were not studied on List 2 (List 1 intact condition). The other six were constructed by making a rearranged pair from items that were only presented on List 1 (List 1 rearranged condition). The participants could correctly classify these foils as new because the items were not on List 2 or because they were not presented together. Table 1 contains an example of each condition. Studied pairs were always presented side-by-side and test pairs were always presented one above the other with no relationship between the study and test position.

Results

A repeated measures analysis of variance (ANOVA) was conducted on the hit rates (HRs) and on the false alarm rates (FARs) with study condition as the within-subjects factor. All ANOVAs are repeated measures and all post hoc tests are Bonferroni adjusted unless otherwise stated. There was a main effect of study condition on the HRs, $F(2, 158) = 25.40$, $MSE = .001$, $p < .001$, and post hoc tests confirmed the order apparent in Figure 1. Namely, the HR was highest for the List 1 and 2 exact condition (M = .680, SEM = .023) followed by Lists 1 and 2 recombined condition (M = .595, SEM = .024), followed by the List 2 condition (M = .520, SEM = .025). There was also a main effect of study condition on the FARs, $F(4, 316) = 16.392$, $MSE = .030$, $p < .001$. FARs in those conditions where items appeared in both lists were higher than the FARs to rearranged pairs constructed from items that appeared on a single list. However, there was no difference in FARs to those foils constructed from repeated items (List 1 and 2 exact, M = .259, SEM = .025; Lists 1 and 2 recombined, M = .277, SEM = .028). Similarly, FARs to those rearranged foils whose items appeared on only one study list, either List 1 only (M = .171, SEM = .021) or List 2 only (M = .184, SEM = .021) did not differ. FARs to intact pairs from List 1 were numerically greater than FARs for any other condition (M = .362, SEM = .026), however according to Bonferroni tests, the List 1 intact FAR differed from all conditions except List 1 and 2 recombined.

In summary, when items in an AR test pair had been presented in the same type of pair in a prior study list, participants were more willing to call the test pair "old" than if the items had been studied only once. The additional tendency to say "old" to pairs containing repeated items was approximately the same for targets (.075) and foils (.093). When test items were presented in an identical pair on both lists, the increase in the HR (.16) was much greater than the increase in the FAR (.082), suggesting that encoding of a pair improves with repetition. Intact foils from List 1 had a very high FAR, indicating a difficulty in list discrimination for AR, as is typical in single item recognition. Experiment 2 contrasts these findings to the case where items are repeated in a different type of pair.

Table 1
An Example of Each Study and Test Condition for Experiment 1

Study List 1	Study List 2	Test pair	Condition label
	8 hue	8 hue (target)	List 2
	9 plea	9 hue (foil)	
3 toil	3 myth	3 myth (target)	Lists 1 and 2 recombined
4 myth	7 toil	7 myth (foil)	
7 watt			
11 bonus	11 bonus	11 bonus (target)	Lists 1 and 2 exact
12 fare	12 fare	12 bonus (foil)	
5 assist		5 civil (foil)	List 1 rearranged
6 civil			
2 cache		2 cache (foil)	List 1 intact

Note. Numbers refer to faces in the actual experiment. In the actual experiment, no item would be repeated during test (as is illustrated here simply to conserve space).

Experiment 2

Method

Participants. Fifty-eight people from the Indiana University community participated in the experiment in exchange for partial course credit or $7.00 per hour.

Materials. The materials were identical to those of Experiment 1.

Procedure. The procedure was identical to that of Experiment 1.

[2] The construction of the Lists 1 and 2 recombined study pairs differed between subjects for both Experiments 1 and 2. Assume the pairs AB, CD, EF, GH, and so forth were studied in List 1. For one group, two resulting recombined pairs would be AD and CB. That is, both items from two studied pairs in List 1 were recombined to form two studied pairs in List 2. For the other group, an item from one studied pair could be paired with any item from another pair except that there were no cases of type of pairing described above. There was no main effect of the type of recombined pair and this variable did not interact with any other variables in either experiment. Thus, the data are presented collapsed over this variable. Because we found no difference between the two methods for recombining study pairs, only the method of random selection was used for Experiments 3 and 4.

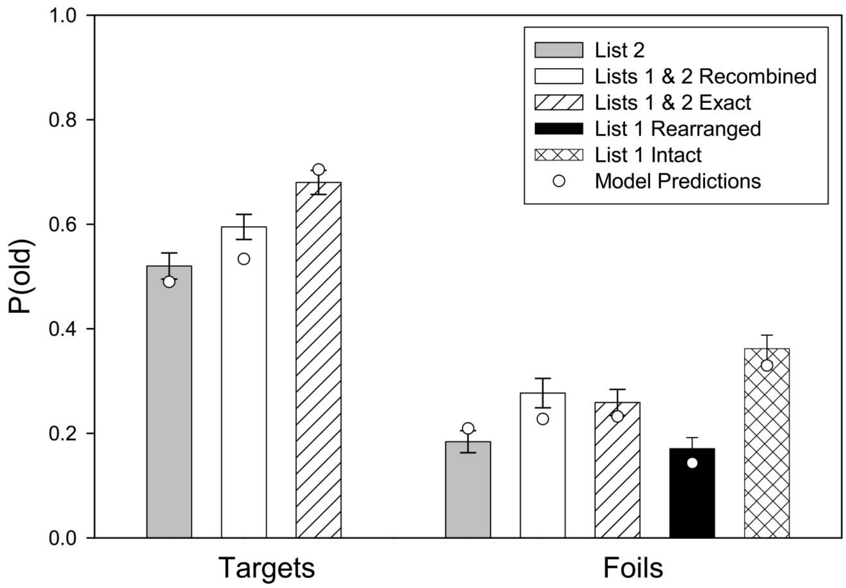

Figure 1. The probability of calling a test item old [P(old)] as a function of the type of test pair in Experiment 1 where item repetitions occurred in the same type of pair. Error bars represent one standard error above and one below the mean. Open circles represent the fit of the REM model described in the text.

Design. List 1 contained 26 WW and 26 FF pairs. The 60 List 2 pairs, all WF, contained a subset of those conditions found in Experiment 1 because of the constraint that a pair could not be repeated exactly on List 2, given that List 2 did not contain the same type of pairs as List 1. When an item was repeated, it occurred in a different type of pair that the initial presentation; this is referred to as the *different* condition. An equal number of List 2 study pairs came from each of the following conditions: studied only on List 2 and items studied on Lists 1 and 2 but in different pairs (Lists 1 and 2 recombined condition). The test list contained 18 intact pairs (targets) and 12 rearranged pairs (foils) from the two conditions described above. In addition, 12 foils were constructed by making a rearranged pair from items that were only presented on List 1 (List 1 rearranged condition). These foils could be called "new" either because the individual items were not presented on List 2 or because the items were not presented together. Table 2 contains an example of each condition.

Results

HRs and FARs are pictured in Figure 2. The HRs for the List 2 condition (M = .533, SEM = .028) and Lists 1 and 2 recombined conditions (M = .570, SEM = .029) did not differ, $F(1, 56)$ = 2.649, MSE = .015, p = .109. FARs differed by study condition, $F(2, 112)$ = 10.701, MSE = .015, $p < .001$. Both conditions that contained items from List 2 had similar FARs (for List 1 and 2 recombined, M = .228, SEM = .023; for List 2, M = .222, SEM = .020) and they were both greater than the FAR to List 1 rearranged foils (M = .137, SEM = .018).

Table 2
An Example of Each Study and Test Condition for Experiment 2

Study List 1	Study List 2	Test pair	Condition label
	7 taper	7 taper (target)	List 2
	8 crisis	8 taper (foil)	
1 2	1 deed	1 deed (target)	Lists 1 and 2 recombined
3 4	3 pious	3 deed (foil)	
lessen deed			
outset pious			
5 6		5 array (foil)	List 1 rearranged
neutral array			

Note. Numbers refer to faces in the actual experiment. In the actual experiment no item would be repeated during test (as is illustrated here simply to conserve space).

Comparison of Experiments 1 and 2

We have noted that when single items are repeated in the same type of pair, participants are more willing to call the resulting test pairs "old" regardless of their actual status (i.e., Experiment 1). However, when item repetitions occur in a different type of pair, we see little to no contribution of the repetitions (i.e., Experiment 2). To draw stronger conclusions about this interaction, we now directly compare the corresponding conditions of the two experiments. A 2 × 2 × 2 mixed designs ANOVA was computed with experimental group as the between-subjects factor and condition (List 1 and 2 recombined and List 2) and test type (target or foil) as the within-subject factors. To confirm the individual analyses, we should find an interaction between experimental group and condition such that the probability of calling and item "old" [P(old)] is greater for the List 1 and 2 recombined condition relative to the List 2 condition when the items repeat in the same type of pair (Experiment 1) but not when items repetitions occur in a different pair type (Experiment 2). Indeed, we do find this interaction between experimental group and condition, $F(1, 137)$ = 9.803, MSE = .022, p = .002. In addition, we find main effects such that P(old) was higher to targets than foils and to the List 1 and 2 recombined condition than the List 2 condition: $F(1, 137)$ = 306.999, MSE = .048, $p < .001$, and $F(1, 137)$ = 16.458, MSE = .022, $p < .001$, respectively. No other interactions were significant, nor was there a main effect of experimental group (all Fs < 1 and ps > .334).

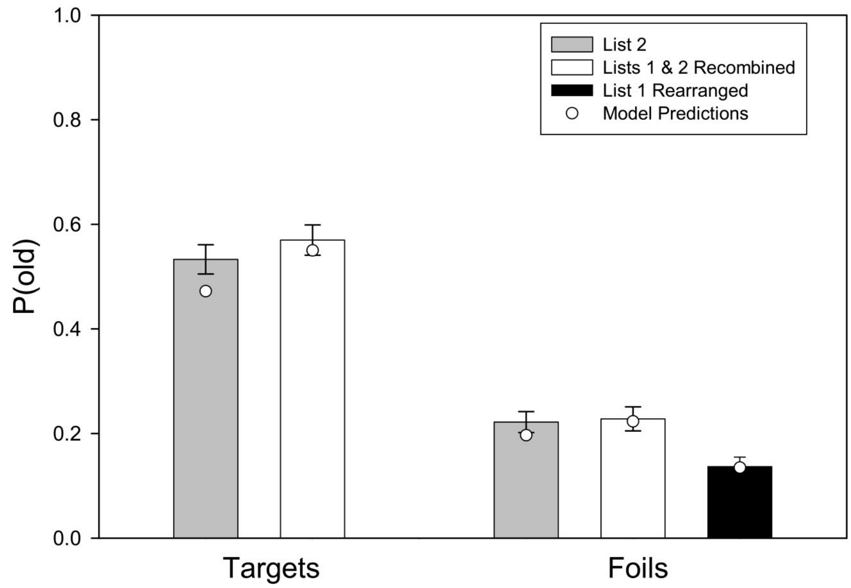

Figure 2. The probability of calling a test item old [P(old)] as a function of the type of test pair in Experiment 2 where item repetitions occurred in a different type of pair. Error bars represent one standard error above and one below the mean. Open circles represent the fit of the REM model described in the text.

Discussion of Experiments 1 and 2

The important difference between the *same* condition (Experiments 1) and the *different* condition (Experiment 2) is the type of pair presented on List 1. In the *same* condition, both Lists 1 and 2 contained the same type of pairs (i.e., WF pairs). In the *different* condition List 1 contained WW and FF pairs, whereas List 2 contained WF pairs. For both targets and foils, seeing the individual items on a previous list in the same type of pair enticed participants to call the test pair "old" more often compared with the case in which items were presented only on List 2. When items were repeated in a different type of pair, it was almost as if List 1 never occurred, as we saw little change in performance. This is consistent with our previous findings showing a list length effect within, but not between, pair types and with models where different pair types are coded with dissimilar and nonoverlapping representations. The tendency to call a pair "old" more often if the items comprising the pair have been repeated is not without precedent. Dyne, Humphreys, Bain and Pike (1990) had participants study a single list of pairs where items repeated between pairs (all WW pairs). Like us, they found an increase in P(old) for AR (they also report no change in d-prime). They showed that a number of models predicted this pattern because pairs with repeated items are more familiar, the increase in familiarity being the same for intact target and rearranged foil pairs. This finding appears to be limited to the case in which item repetitions occur in the same type of pair.

Experiments 3 and 4

In the models proposed in Criss and Shiffrin (2004b), we assumed that single item and pair representations differed. This was based in part on two empirical findings: Performance in single-item recognition was determined by the total length of the list and not the relative number of pairs of each type, and the overall level of accuracy for the different pair types did not predict the level of accuracy for single items. However, in the original discussion of these issues, we did not specify the relationship between the features identifying the pairs and those identifying the singles from which they were constructed. One could imagine a model in which the pair type biases the encoding of the single item to include those features relevant to the studied pair. For example, previous studies have shown that the encoding of the word *jam* is different when studied in the pair *strawberry jam* than when studied in the pair *traffic jam* indicating an item encoding that is specific to the pair in which it was studied, at least when the items are related pre-experimentally (e.g., Light & Carter-Sobell, 1970; Tulving & Thompson, 1973).

Strengthening some study items via spaced repetition typically results in a different pattern of performance for free recall and recognition. A positive list-strength effect, defined as a decrement in performance for the nonstrengthened items compared with a list in which all items are of equal strength, occurs for free recall but a null or negative list-strength effect occurs for single-item recognition (Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Ratcliff, & Clark, 1990). This was modeled by assuming that item repetitions result in differentiation, thereby reducing interference (i.e., strengthening items makes them less similar to other items). Murnane and Shiffrin (1991) asked if differentiation of single items occurs when items were repeated within sentences. In one condition, whole sentences repeated. In another, each individual sentence occurred just once, but each sentence was constructed from words that repeated in other sentences. There were various testing conditions but single item recognition proves to be the most relevant for the current discussion. When strengthening was accomplished by repeating a whole sentence, a null or negative list-strength effect was found, replicating earlier results in which single items were studied. On the other hand, when strengthening was accomplished by rearranging the repeated words into different sentences, a positive list-strength effect was obtained (even though the same words were repeated the same number of times for each of the conditions described). In the resulting model, they assumed that the encoding of individual words is biased by its surroundings such that a word repeated in three different sentences acts functionally as three different words, preventing differentiation. Apparently, a word repeated in the same sentence acts as the same word each time, producing differentiation and making that stored word less similar to other studied words.

The two studies just described assume that the encoding of a single item is biased by the surrounding context. Such explanations may be contrasted with models assuming an emergent set of features for associations and a relatively stable representation for singletons. These models allow different qualitative patterns of prediction for tests involving single items and some combination of items because the tasks are based on different sets of features (associative or single item) containing different information. The results of the following studies allow us to better understand which of these assumptions is most appropriate as we continue to develop the model proposed in Criss and Shiffrin (2004b) and described later in this article. The study conditions of these experiments are identical to Experiments 1 and 2, but participants are given an unexpected single-item recognition test following study. In Experiment 3, single items are repeated in the same type of pair across lists, and in Experiment 4 single items are repeated in a different type of pair.

Experiment 3

Method

Participants. Twenty-five people from the Indiana University community participated in the experiment in exchange for partial course credit or $7.00 per hour.

Materials. The materials were identical to those of Experiment 1.

Procedure. Participants received two study lists separated by an unfilled break of at least 120 s. The first study list contained 52 pairs of items and the second contained 60 pairs. On each trial of each list, participants had 3 s during which they performed an incidental task that involved rating each pair on the following question: "Do these two items go together?" Each study trial was separated by a 500-ms interstimulus interval. Following the final study list, participants were engaged in a 45-s math task before beginning an unexpected memory test consisting of 120 single items presented one at a time. Participants were instructed to respond with "old" only if the single item had been studied on List 2.

Design. The study lists were constructed just as those in Experiment 1. The test list consisted of 120 trials, half words and half faces. The targets consisted of an equal number of words and faces from each of the three study conditions: List 2, Lists 1 and 2 recombined, and Lists 1 and 2 exact. The foils consisted of six faces from List 1, six words from List 1, and 48 items (half faces and half words) that were not previously studied.

Results

A 2 × 3 (item type and study condition) ANOVA was conducted on the HRs. HRs were higher for faces than words, $F(1, 24) = 5.932$, $MSE = .042$, $p = .023$. There was a main effect of study condition, $F(2, 48) = 15.880$, $MSE = .026$, $p < .001$, and no interaction between the two, $F(2, 48) = 0.092$, $MSE = .020$, $p = .912$. Post hoc analyses confirm what is shown in Figure 3, Panel A—namely, that HRs for items presented on Lists 1 and 2 ($M = .606$, $SEM = .037$, for the exact condition, and $M = .642$, $SEM = .037$, for the recombined condition) are both greater than the HR for items presented only on List 2 ($M = .470$, $SEM = .035$) but do not differ from one another.

A 2 × 2 (item type and foil type) ANOVA conducted on the FARs showed higher FARs for faces than words, $F(1, 24) = 7.236$, $MSE = .027$, $p = .013$. False alarms to those items presented on List 1 ($M = .280$, $SEM = .029$) were much higher than false alarms to new items ($M = .111$, $SEM = .016$), $F(1, 24) = 26.504$, $MSE = .027$, $p < .001$, and there was no interaction between item type and foil type $F(1, 24) = 0.082$, $MSE = .026$, $p = .777$. Figure 3, Panel A shows the hits and false alarms collapsed over item type. For a breakdown by item type, see Table 3.

Experiment 4

Method

Participants. Twenty-five people from the Indiana University community participated in the experiment in exchange for partial course credit or $7.00 per hour.

Materials. The materials were identical to those of Experiment 1.

Procedure. The procedure was identical to that of Experiment 3.

Design. The study lists were constructed just as those in Experiment 2. The test list consisted of 120 trials, half words and half faces. The targets consisted of an equal number of words and faces from the List 2 and Lists 1 and 2 recombined conditions. The foils consisted of six faces from List 1, six words from List 1, and 48 items (half faces and half words) that were not previously studied.

Results

A 2 × 2 (item type and study condition) ANOVA was conducted on the HRs. The HR to items presented on both Lists 1 and 2 ($M = .667$, $SEM = .035$) was much higher than the HR to items presented only on List 2 ($M = .535$, $SEM = .031$), $F(1, 24) = 22.246$, $MSE = .020$, $p < .001$. There was no main effect of item type and no interaction between the two variables, $F(1, 24) = 1.090$, $MSE = .030$, $p = .307$, and $F(1, 24) = 0.189$, $MSE = .012$, $p = .668$, respectively).

A 2 × 2 (item type and foil type) ANOVA was conducted on the FARs. FARs to items presented on List 1 ($M = .353$, $SEM = .044$) were greater than the FARs to new items ($M = .120$, $SEM = .022$), $F(1, 24) = 42.86$, $MSE = .032$, $p < .001$. There was no difference between FARs to words and faces though the effect approached statistical significance, $F(1, 24) = 3.349$, $MSE = .019$, $p = .080$, and there was no interaction between the two variables, $F(1, 24) = 0.729$, $MSE = .019$, $p = .402$. Figure 3, Panel B shows the hit and false alarms collapsed over item type. For a breakdown by item type, see Table 4.

Comparison of Experiments 3 and 4

Observation of Figure 3 along with the individual statistical analyses from Experiments 3 and 4 both indicate the same pattern of results for single item recognition regardless of whether the single items are repeated in the same or different type of pair. Here we directly compare the corresponding conditions of the two

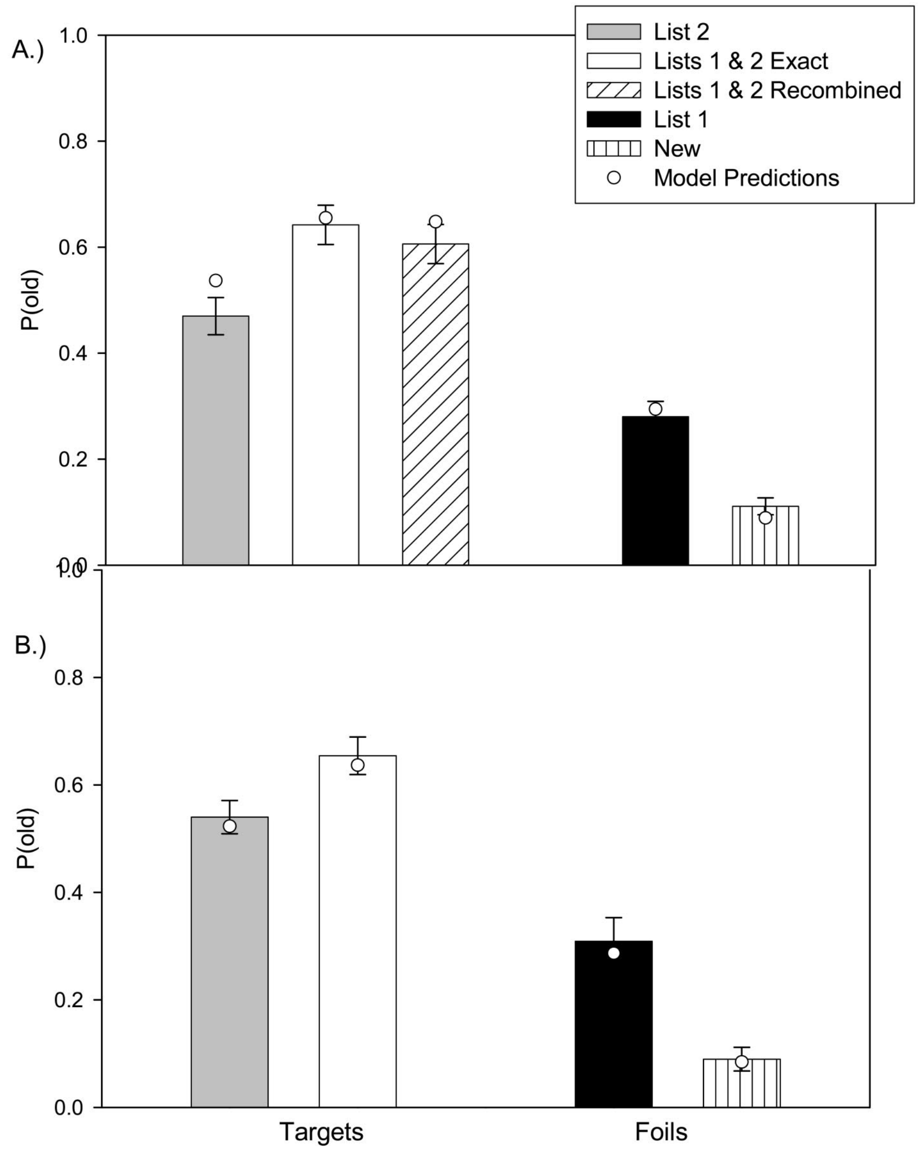

Figure 3. The probability of calling a test item old [P(old)]as a function of the type of test item. Panel A shows the data from Experiment 3 (repetitions in the same pair type) and Panel B shows the data from Experiment 4 (repetitions in a different pair type). Error bars represent one standard error above and one below the mean. Open circles represent the fit of the REM model described in the text.

experiments. A 2 × 2 × 2 mixed designs ANOVA was computed for the HR and the FAR. In both cases, experimental group was the between-subjects factor and condition (List 2 and List 1 and 2 recombined) and item type (faces and words) were the within-subject factors. Those items studied on both lists have a higher HR than those studied only on List 2, $F(1, 48) = 43.477$, $MSE = .027$, $p < .001$). We found an interaction between item type and experiment due to the higher HR for faces than words in Experiment 3 but not in Experiment 4. Given that the total number of studied faces and words is equal for the two groups, there is no obvious reason for this pattern of data, and it is simply attributed to idiosyncratic difference between groups of participants. There were no other main effects or interactions for targets (all Fs $<$ 1.097 and all ps $>$ .300). Words have lower FARs than faces, $F(1, 48) = 10.494$, $MSE = .023$, $p = .002$, and items that were never studied have lower FARs than items presented on the first list, $F(1, 48) = 68.940$, $MSE = .029$, $p < .001$. There were no other main effects or interactions for foils (all Fs $<$ 1.752 and all ps $>$ .192). Thus, as expected given the individual analyses, we found the same pattern of data for single-item recognition regardless of whether item repetitions occurs in the same or different type of pair. In particular, we found that targets presented on both lists were more likely to be called "old" than were targets studied only on the second list. Likewise, foils studied on the first list were more likely to be called "old" than were foils that were never studied during the experiment.

Table 3
Hit Rates and False Alarm Rates as a Function of Item Type for Single Item Recognition in Experiment 3

Condition	Faces	Words
Hit rate		
List 2	.504 (.043)	.436 (.041)
Lists 1 and 2 recombined	.688 (.042)	.596 (.044)
Lists 1 and 2 exact	.648 (.040)	.564 (.049)
False alarm rate		
List 1	.320 (.047)	.240 (.036)
New	.160 (.025)	.062 (.014)

Note. Standard errors of the mean are listed in parentheses.

Discussion of Experiments 3 and 4

In summary, we found the same pattern of data for single-item testing regardless of whether items were repeated in the same or different type of pair. This result contrasts with the pattern found for associative recognition, in which repetitions of the same type shape performance, suggesting that single items are stored in a similar fashion regardless of the type of pair in which they were encoded. The current conclusions also differ from those of Murnane and Shiffrin (1991), who showed a positive list strength effect for single words repeated in different sentences but a typical null list strength effect for words repeated in the same sentences. They proposed that a word studied in different sentences acts as a different word each time, with a meaning biased by the sentence context. They noted an unpublished study, using pairs of words rather than five-word sentences, that failed to find a similar pattern. Murnane and Shiffrin suggested that a sentence places more constraints on the meaning of the constituent words than a pair, thus biasing the encoding in a stronger way than study of a pair. Thus, though our results differ from that of Murnane and Shiffrin, the differences may be due to the differing constraints imposed by sentence contexts versus pair contexts. We reserve further tests of this hypothesis for future studies.

A REM Model for Three-Phase Associative and Single-Item Recognition

We now describe the model first suggested by Criss and Shiffrin (2004b), and show that this model successfully predicts the patterns of data found in Experiments 1–4 both qualitatively and quantitatively. Criss and Shiffrin (2004b) proposed a representational schema that could be added to nearly any extant model to accommodate different classes of pairs. In the following model, we implemented one of those representations within the REM model which was first proposed for single item and associative recognition by Shiffrin and Steyvers (1997, 1998). Further, we specified the relationship between associative features and the single items from which they were created. Finally, we also implemented an assumption suggested in Criss and Shiffrin (2004b) to account for the Kelley and Wixted (2001) data and used it to successfully account for list discrimination in AR.

According to the REM model, each study trial results in the storage of a memory trace that contains several sets of features. In the original conception, memory traces contained features describing the item and features describing the current environmental and internal context. Study of a pair simply resulted in two sets of item features stored in the same memory trace. However, this concatenation assumption was challenged by the set of data described here and in Criss and Shiffrin (2004b). Such results raise the possibility that encoding of WF, FF, and WW pairs results in dissimilar associative information, despite the fact that pairs might share single items of the same type. We propose that this associative information manifests in two ways. First, following Murdock (1982), we allow the storage of associative features, a unique set of features generated on the fly for each pair, that are independent of single item features (cf. Clark & Gronlund, 1996; Dosher & Rosedale, 1997; Hockley & Cristi, 1996b; Kahana, 2002). Under this assumption, the associative features resulting from AB are no more similar to AD than to EF, despite the shared single item. Likewise, the single item features for A are only similar to the associative features of AB by chance. This assumption allows qualitatively different patterns of results for singles and pairs, but does not produce the functional independence of the three different types of pairs. Second, we allow for a set of type code features so that each class of pairs can be accessed more or less separately from the others. We consider the type code to be analogous to a category cue, a class attribute (cf. Galbraith, 1975; Underwood, 1969), or set designating features (cf. Shiffrin & Steyvers, 1997). As Underwood illustrated, when attempting to generate a technical term, one does not generate the name of a colleague, and when recalling CVC strings, one does not generate a known five-letter word. A more empirically based example of such a process is found in studies of the fan effect in which participants were asked to verify facts about different categories. Response time in these tasks depends on the number of categories and the number of facts within the category being tested, but not the number of facts learned about irrelevant categories (Anderson & Paulson, 1978; McCloskey & Bigler, 1980; Reder & Anderson, 1980). Thus, intuitively and empirically, there is evidence that participants are able to limit the memory search to a particular subset of memory given sufficient cues. We assume that the type code is such a cue. Though Underwood assumed this type of cue could be used strategically by the participant or when instructed by the experimenter, we assume the type code is used when available. For example, during AR the type of pair is obvious, thus the type code is used. However, in other tasks in which the probe consists of only a single item, such as single item recognition or free recall, the type of pair that is relevant for the task is not obvious, thus it is not adaptive to use a

Table 4
Hit Rates and False Alarm Rates as a Function of Item Type for Single Item Recognition in Experiment 4

Condition	Faces	Words
Hit rate		
List 2	.512 (.039)	.557 (.037)
Lists 1 and 2 recombined	.653 (.040)	.680 (.037)
False alarm rate		
List 1	.367 (.050)	.340 (.050)
New	.157 (.029)	.083 (.020)

Note. Standard errors of the mean are listed in parentheses.

type code on the basis of pair type. Whether or not participants can use a type code when explicitly instructed to do so is a topic left for future research.

As just described, study of a pair under instructions emphasizing associative encoding results in the storage of item, context, associative, and type code features. Whether or not these features contribute to the memory decision depends on the type of test employed. For example, when probed with a single item (as in Experiments 3 and 4 or in a recall task), the type code and associative features are not available and thus presumably do not contribute to the retrieval and decision process. Thus, single item recognition proceeds just as described in Criss and Shiffrin (2004a): Item and context features of the test probe are compared with the same type of features in each memory trace, resulting in a likelihood value that each memory trace resulted from study of the test probe. The likelihoods are combined into a single value indicating the overall familiarity of the test probe and a decision is based on that value. The fit to the set of single item data shown in Figure 3 required no modification to the model.

Now consider associative recognition. AR differs from single item recognition in the types of probes used during retrieval. Following Murdock (1997), we assume context features are not an effective cue for an AR task. We therefore assume these are not used. Likewise, for the following reason, the familiarity of the single items in the test probe is ignored: In a standard AR paradigm, using single-item familiarity as a basis for a memory decision can only harm performance, as all individual items were studied and are familiar. Thus, in our AR model the retrieval processes and memory decision are based on the type code features used to restrict comparisons to the relevant subset of memory and on the match between stored associative features and those generated from the pair of test items.

However, there are AR studies in which consideration of single-item familiarity is adaptive. In Criss and Shiffrin (2004b), we pointed out that under certain conditions, participants may adopt a strategy of using single item familiarity to help reject foils. Specifically, we suggested that such a strategy may be used when context information and/or single item familiarity is useful for the task, such as in Kelley and Wixted (2001). In their paradigm, participants were tested with intact and rearranged test pairs as well as foils constructed from two unstudied items. They found that the FAR for unstudied foils fell below the FAR to rearranged foils (among other manipulations and findings). We suggested that their participants used single item familiarity to augment their AR decisions as follows: If both singles were judged to be new then the pair would be called "new," otherwise the judgment would be based on associative features alone. We assume participants adopted the same retrieval strategy to perform list discrimination in the current experiments. In summary and as illustrated in Figure 4, we assumed that the initial probe with type code features restricts further comparisons to a relevant subset of memory. Then single item and associative features are compared with those traces in this activated set. If both singles are rejected, then the pair is called "new," otherwise a decision is based on the match between

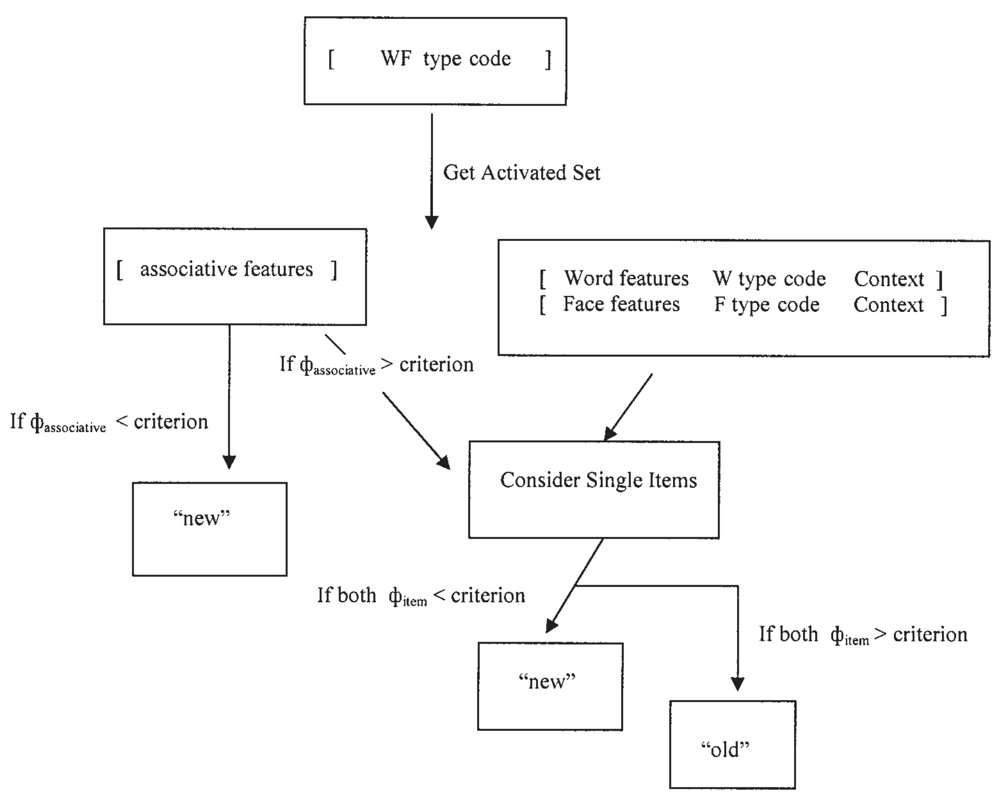

Figure 4. A schematic of the retrieval processes involved in the model used to generate the fits depicted in Figures 1, 2, and 3. See the text for a detailed description of the model. W = word; F = face.

the associative features of the probe and those stored in memory traces in the activated set.

The model predictions, shown as white circles in Figures 1, 2, and 3, are quite close to the observations with no major deviations, despite the limited parameter search and the constraint that all parameter values excluding one were identical for all four groups. In a likelihood based model such as REM, the optimal criterion is defined as the point of indifference where the test probe is equally likely to have come from a studied item as it is to have been an unstudied item. To maximize percent correct, one should call all items with an odds value exceeding this criterion "old" and all those with a value less than this criterion "new." To accurately fit the current set of data, we needed to use nonoptimal criteria. In particular, the criterion for saying "old" is more strict than optimal for single-item testing and more lenient than optimal for AR. Given that the similarity between the two study lists (and hence the similarity of the context features) boosts the familiarity of all items studied on either list, it makes sense that participants would be conservative in claiming an item was on the second study list. On the other hand, some tested pairs will be rejected on the basis of the familiarity of the single items (regardless of the familiarity of the associative features), thus it seems reasonable that participants may be slightly liberal in calling a test pair "old" on the basis of associative features.

Further details of the model can be found in the original sources, and details for the simulations discussed here can be found in the Appendix. Some readers may have the impression that the assumptions underlying our present model are ad hoc and/or post hoc. It is noteworthy, therefore, that the model used here was suggested in Criss and Shiffrin (2004b) before the present data were collected as a plausible way to accommodate both the within-class but not between-class list length effects and Kelley and Wixted's (2001) data. This model is applied here to a quite different paradigm and yet fits with high accuracy. Good fit notwithstanding, we admit that variants of the specifics of this model are possible. We suspect that all such variants would have to incorporate some form of dissimilar representation for different pair types. For example, recall the persistence in encoding hypothesis prevalent in the paired associate literature, according to which a repeated item tends to be encoded in a way that is consistent with its last encounter. In the present paradigm, one might assume that persistence in encoding only occurs when an item is repeated in the same type of pair. Such an assumption would attribute our findings to a mechanism during encoding rather than a retrieval strategy as adopted in the current model.

Studies measuring receiver operating characteristics and decision time suggest that AR may be carried out via a search process (e.g., Nobel & Shiffrin, 2001; Rotello & Heit, 2000; Rotello, Macmillan, & VanTassel, 2000; Yonelinas, 1997; but see Gronlund & Ratcliff, 1989). Why then did we adopt a familiarity-based model? In part, we did so because the present results did not allow us to distinguish a familiarity-based decision from one based on an elaborative search process, thus we chose to implement the simpler familiarity process. However, some details of the data suggest that a strategy of successively using each individual test item to recall its studied partner is not the underlying process. For example, one might expect a recall model to predict lower HRs for the Lists 1 and 2 recombined condition compared with the List 2 condition because of competition between the two study partners in the former case. In other words, a recall-to-accept strategy should be less successful for the Lists 1 and 2 recombined condition because of interference. Similarly, one might expect a lower FAR for the List 1 and 2 exact condition because of a greater ability to recall the correct study partner given that the exact pair was studied twice compared with either condition in which pairs were studied just once (e.g., List 2) or studied twice with different partners (e.g., Lists 1 and 2 recombined). That is, a recall-to-reject strategy should be relatively more successful for the Lists 1 and 2 exact condition. Of course, the exact predictions depend on the particular model implementation. Critically, any such search model would require different representational similarity for different pair types to account for the interaction between the *same* versus *different* AR conditions This idea is, of course, the main point of this article.

We should also consider further the assumption that context plays no role in AR. Though supported by studies showing no forgetting for pairs relative to singles over a moderate range of study–test lags (Hockley, 1992; Hockley & Consoli, 1999), this assumption sounds a bit extreme. It seems likely that context features are part of the AR probe but play a less important role for various reasons including limited capacity and lack of necessity. This line of thinking does not imply, however, that one can simply generate an alternative to the present model in which context features are used in the associative probe as a replacement for the assumption that single-item familiarity is used in decisions. Recall that representations for a pair and its constituent single items are assumed to be independently generated; thus, the familiarity of the single items does not affect the familiarity of the associative features. For this reason, an alternative model of the type just described predicts an approximately equal FAR for rearranged pairs regardless of the number of times or lists on which the individual single items were studied. Because the context features used in a probe are identical on each test trial, they add a constant amount of evidence to the decision regardless of the type of rearranged foil. For the same reason, this alternative model predicts an approximately equivalent HR for all pairs that were presented only on List 2 (i.e., the List 2 and Lists 1 and 2 recombined conditions), regardless of whether the single items comprising the pair were studied on the first list. These predictions do not agree with our data. Thus, though some readers may think it logical and intuitive to include context features in the associative probe, it would not allow the model to better predict the observed pattern of data.

Thus, to predict our data, we were led to the incorporation of single item familiarity. For example, whereas the Lists 1 and 2 exact condition has a higher HR than the other conditions because of the inclusion of associative features from both study lists, the List 2 HR is lower than the Lists 1 and 2 recombined HR because of the occasions on which both single items in the List 2 condition are determined to be new and the pair is rejected. This outcome is more likely for List 2 pairs than Lists 1 and 2 recombined (or exact) pairs because each item was studied only once in the former case. Similarly, the FARs to pairs containing items studied once (i.e., List 1 and List 2) are lower than FARs to pairs containing items studied twice (i.e., Lists 1 and 2 exact and recombined) because of the once presented items being rejected more often than twice presented items. The use of context in an associative probe would primarily serve to reduce the FAR to intact pairs studied on the first list, allowing a concrete test of the assumption. Suppose

List 1 contains pair AD and List 2 contains pairs AB and CD. The test pair AD (given the same instructions used here, to say "old" only to List 2 pairs) should have a FAR approximately equal to the HR of an intact pair studied on List 2. According to the model, if both singles are judged "new" then the pair is rejected, otherwise the decision is based on the associative features. In this case, both single items A and D were studied on List 2, so the decision will likely be based on the associative features. The pair AD was studied on List 1 and, without using context features in the probe, the associative features alone will likely result in an "old" decision.

In summary, no extant models are able to account for the results of this study and the data presented in Criss and Shiffrin (2004b) without additional assumptions. The co-occurrence models cannot account for the current set of data because we find a qualitatively different pattern of results for AR and single item recognition. The emergent feature assumption of composite models has been supported here and adopted in our own model. However, composite models in the form that they exist presently cannot handle our data. These models combine all memory traces into a single vector, causing all memory traces to contribute to the decision for each other. Because we find larger effects in AR when the item repetitions occur in the same type of pair and list length effects restricted to pairs of the same type, composite models would also require an assumption that similarity differs between different pair types. This is accomplished in the present model by adopting a type code. Without the type code, all associative features would participate in the decision and the model would predict a similar pattern of data regardless of whether repetitions occurred in the same or different type of pair. Also, some composite models (e.g., TODAM) share the assumption that context is not part of the probe for AR testing and thus predict approximately equivalent FAR for all rearranged foil types and approximately equivalent HR for those intact pairs studied once. Though we do not consider all possible search models, the details of the data suggest that a recall-to-reject or a recall-to-accept strategy using single items as probes is not sufficient. Thus, it seems that to account for the current set of data and the list length findings of Criss and Shiffrin (2004b), extant models require mechanisms allowing pairs to be selectively involved in the decision process depending on the type of pair being tested and to allow differential interference for probes of pairs compared with singles. We used type codes and emergent associative features in the REM framework as an example of how one could extend a model to fit this set of data.

Summary

In a paradigm in which items were studied on multiple lists, we have shown that changes in memory performance on a subsequent associative recognition task depend on the type of pair in which the repeated items were studied. We found a large increase in P(old) for both targets and foils when items were repeated across lists in the same type of pair, but not when repetitions occurred in different types of pairs. In contrast, performance on a single item recognition task is not subject to such pair-type dependencies. This data was well fit by a model assuming that participants adopt a strategy of using single-item familiarity to help make a list discrimination decision in AR. In particular, if the stored features for both single items indicate that neither was studied on the relevant study list, the pair is called "new" regardless of the familiarity of the associative features. Otherwise, the AR decision is based strictly on the familiarity of the associative features.

References

American Telephone & Telegraph. (1994). The Olivetti research database of faces [electronic database]. Available from (FTP: Hostname) Uk .research.att.com Directory: Pub/data/att_faces.tar.Z.

Anderson, J. R., & Paulson, R. (1978). Interference in memory for pictorial information. *Cognitive Psychology, 10,* 178–202.

Asch, S. E. (1964). The process of free recall. In C. Scheerer (Ed.), *Cognition: Theory, research, promise* (pp. 79–88). New York: Harper & Row.

Asch, S. E. (1969). A reformulation of the problem of associations. *American Psychologist, 24,* 92–102.

Begg, I. (1978). Imagery and organization in memory: Instructional effects. *Memory & Cognition, 6,* 174–183.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

Clark, S. E., Hori, A., & Callan, D. E. (1993). Forced-choice associative recognition: Implications for global-memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 871–881.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505.

Criss, A. H. (2005). The representation of single items and associations in episodic memory (Doctoral dissertation, Indiana University, 2004). *Dissertation Abstracts International-B, 65,* 6882

Criss, A. H., & Shiffrin, R. M. (2004a). Context noise and item noise jointly determine recognition memory: A comment on Dennis & Humphreys (2001). *Psychological Review, 111,* 800–807.

Criss, A. H. & Shiffrin, R. M. (2004b). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition, 32,* 1284–1297.

Dosher, B. A., & Rosedale, G. S. (1997). Configural processing in memory retrieval: Multiple cues and ensemble representations. *Cognitive Psychology, 33,* 209–265.

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,* 813–824.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62,* 145–154.

Galbraith, R. C. (1975). On the independence of attributes of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1,* 23–30.

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

Greeno, J. G., James, C. T., DaPolito, F., & Polson, F. G. (1978). *Associative learning: A cognitive analysis.* Englewood Cliffs, NJ: Prentice Hall.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Memory & Cognition, 15,* 846–858.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 94,* 528–551.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition, 26,* 449–462.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1321–1330.

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition, 27,* 657–664.

Hockley, W. E., & Cristi, C. (1996a). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24,* 202–216.

Hockley, W. E., & Cristi, C. (1996b). Tests of the separate retrieval of item and associative information using a frequency-judgment task. *Memory & Cognition, 24,* 796–811.

Howard, M. W., & Kahana M. J. (2002). A distributed representation of temporal contact. *Journal of Mathematical Psychology, 46,* 269–299.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

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

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Memory & Cognition, 27,* 701–722.

Klein, K., Criss, A. H., & Shiffrin, R. M. (2004). Recency judgments and list context. *Proceedings of the 26th annual meeting of the Cognitive Science Society* (p. 1578). Mahwah, NJ: Erlbaum.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning & Verbal Behavior, 9,* 1–11.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 539–559.

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review, 75,* 421–441.

McCloskey, M., & Bigler, K. (1980). Focused memory search in fact retrieval. *Memory & Cognition, 8,* 253–264.

McGee, R. (1980). Imagery and recognition memory: The effects of relational organization. *Memory & Cognition, 8,* 394–399.

McGeoch, J. A. (1942). *The psychology of human learning.* Oxford, England: Longmans, Green.

Melton, A. W., & Martin, E. (1972). *Coding processes in human memory.* Oxford, England: Winston & Sons.

Mensink, G. M., & Raaijmakers, J. G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology, 33,* 172–186.

Metcalfe-Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92,* 1–38.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104,* 839–862.

Murdock, B. B., & Kahana, M. J. (1993). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 689–697.

Murnane, K., & R. M. Shiffrin. (1991). Word repetitions in sentence recognition. *Memory & Cognition, 19,* 119–130.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 384–413.

Postman, L. (1976). Interference theory revisited. In J. Brown (Ed.), *Recall and recognition* (pp. 157–182). Oxford, England: Wiley.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Reder, L. M., & Anderson, J. R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology, 12,* 447–472.

Rock, I., & Ceraso, J. (1964). Toward a Cognitive Theory of Associative Learning. In C. Scheerer (Ed.), *Cognition: Theory, research, promise* (pp. 110–146) New York: Harper & Row.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28,* 907–922.

Rotello, C. M., Macmillan, N. A., & VanTassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language, 43,* 67–88.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM–retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73–95). Oxford, England: Oxford University Press.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76,* 559–573.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–763.

Appendix

Details of the REM Model

During each trial of a study list, a memory trace is stored in the form of a matrix. Each memory trace contains features describing the current context, each item, the emergent association between the two items, and the type code indicating what type of pair was studied. REM assumes that features differ in their environmental base rates and thus diagnosticity. This is implemented by independently generating each feature according to a geometric distribution with parameter g as follows:

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

where V refers to the feature being generated and j refers to some specific feature value (j =1, 2, 3, and so forth). We let 15 features represent each part of the vector listed above and set g = .40. Because of imperfect encoding, only some of the available features are stored in the memory trace. In particular, each stimulus feature is stored with some probability, u, otherwise a zero is stored indicating a lack of information. Given that a feature is stored, the correct value is copied with some probability, c = .90, otherwise a random value is drawn from the same geometric distribution and stored. It seemed natural to assume that associative features require more effort than single-item features to generate and to store. We therefore allowed one value of u for associative features, $u_{associative}$ = .20, and another for all remaining features with $u = .32$.

Though context likely changes on a trial by trial basis as a function of random fluctuation (Estes, 1955; Mensink & Raaijmakers, 1989) and/or other studied items (Howard & Kahana, 2001), we use the standard simplifying assumption that context features are fixed during a single study list (cf, Klein, Criss, & Shiffrin, 2004). The current experimental design contains two study lists that are relatively similar both in time and in the encoding task being performed. Thus, we assume the context features are correlated across the two lists. In particular, List 1 context features are selected randomly from the specified geometric distribution. Then List 2 context features are generated by copying each of the List 1 features with some probability, p_{ctx} = .70, and randomly selecting new values from the geometric distribution otherwise.

Individual items are only randomly similar to one another and to their respective association(s). If a study pair contains repeated items, the same set of features are used to store both single item repetitions regardless of pair type. However, the associative features for the two study pairs (assuming the pairs are different as in the Lists 1 and 2 recombined study condition) are generated independently and are similar only by chance. For repetitions of pairs (as in the List 1 and 2 exact study condition) the same associative and the same item features are used during storage. The type code is simply implemented as a set of feature values that is identical for pairs of the same type and only randomly similar to pairs of a different type. One might wish to allow for similarity between type codes (e.g., WF and WW type codes might be similar by virtue of sharing a word). However, in Criss and Shiffrin (2004b), we found a similar lack of interference between all three pair types regardless of the surface similarity and thus do not build in any similarity between type codes. All of the features just described are presumably encoded during the study of a pair under instructions encouraging associative encoding. Whether or not each set of features contributes to the memory decision depends on the type of test employed, as described in the main text. Likewise if the encoding instructions emphasize singletons, fewer associative features (if any) might be stored (e.g., Begg, 1978; Criss, 2005; Hockley & Cristi, 1996a; McGee, 1980).

Single item testing proceeds just as described in Criss and Shiffrin (2004a). Memory is probed with the item features for the current test stimulus and the List 2 context features (because the task requires an “old” response only to List 2 items). These are compared to each stored memory trace, denoted by i, and a matching value is calculated as a likelihood ratio for the match between the probe and each trace. For the item features, a likelihood value is calculated for a memory trace i in the following way:

$$\lambda_{iI} = (1 - c)^{niq} \prod_j \left[\frac{c + (1 - c)g(1 - g)^{j-1}}{g(1 - g)^{j-1}} \right]^{njim}, \tag{2}$$

where niq is the number of nonzero mismatching single item features and $njim$ is the number of matching single item features with the value j. Features stored as zeros are ignored as they represent a lack of information. Because this likelihood is based on item features alone, it is termed $\lambda_{iI,}$ and it gives the degree to which the memory trace matches the probe in item information. In parallel, memory is probed with the relevant context features. Another likelihood value, λ_{iC}, is calculated by comparing the relevant context features (e.g., List 2) with the context features stored in each trace using Equation 2. In this case, niq is the number of nonzero mismatching context features and $njim$ is the number of matching context features with the value j. The term λ_{iC} gives the degree to which the probe matches in context information and is based on context features alone. A recognition test requires that the probe match both item and context information, so the two likelihood values must be combined appropriately. As proposed in Criss and Shiffrin (2004a), we combine the two using a weighting parameter, α, that allows the system to differentially weight item or context information as follows:

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

where ϕ is the odds that the test item was studied in the relevant context and N is the number of memory traces contributing to the decision. If the odds value is greater than some criterion, the item is called “old,” otherwise it is called “new.” If $\alpha = 1$, all decision noise comes from the item features and the context features are ignored and vice versa. As it turns out, a value of α = .50 was used for the current simulations, implying an equal weighting of item and context information. For single item testing, it was necessary to use a criterion of 1.5 indicating that participants were conservative and only claimed an item was studied if it was very familiar. The fit to the set of single item data shown in Figure 3 required no modification to the original model. In our simulations, the number of features and the values of c, g, and α are fixed as specified above. The values of u, p_{ctx}, and the criterion were adjusted to produce a good fit to the overall level of performance.

Figure 4 illustrates the decision process in an AR task when the test probe is a WF pair. First, the type code features are used to probe memory, and a likelihood value for each memory trace is computed via Equation 2 where niq is the number of nonzero mismatching type code features and $njim$ is the number of matching type code features with the value j. Any trace with a likelihood value equal to or greater than some threshold, τ = 1, are considered part of the activated set and are compared with the remaining probes. Note that because of error at encoding, it is possible for pairs of a different type to be mistakenly included in and pairs of the same type to be excluded from the activated set.

Next, associative features for the test pair are generated and compared with the associative features stored in traces contained in the activated set. The comparisons are computed via Equation 2, where niq is the number of nonzero mismatching associative features and $njim$ is the number of matching associative features with the value j. The resulting likelihood ratios, $\lambda_{iA,}$ give the degree to which each memory trace matches the test probe in associative features. The associative activations are combined by the following equation into an odds, $\phi_{\text{associative}}$:

(*Appendix continues*)

$$\phi_{\text{associative}} = \frac{1}{N}\sum_{i}\lambda_{iA}. \quad (4)$$

Even for associative tests, we assume that the familiarity of single items is automatically computed using Equations 2 and 3, resulting in an odds value, ϕ_{item}, for each individual item. (Note that the familiarity of the single item is based on the match between the probe and those traces in the activated set. Hence, not all memory traces are included. In particular, the activated set tends to contain traces that resulted from study of the type of pair being tested—WF in the example in Figure 4.) As discussed in the main text, the single item familiarity is only used when adaptive given the task at hand, such as the case here. Assume therefore that participants given a list discrimination AR test follow the retrieval route suggested above and have available for decisions $\phi_{\text{associative}}$ for the test pair and ϕ_{item} for each individual item in the test pair. If both single items are judged to be new (using for each a default optimal criterion of one for the odds), the pair is called "new." If either of the items is judged to be "old," the AR decision is completely determined by $\phi_{\text{associative}}$. We adopted a criterion of 0.9, indicating that participants were somewhat generous in calling pairs "old," perhaps sensible given that most pairs containing items from List 1 were presumably rejected based on the single item odds.

No parameters were allowed to vary between the *same* and *different* groups. The criteria varied between the groups tested with single items or pairs and the probability of encoding a feature (i.e., u) was lower for associative features than all other features. Fits were not completely optimized, but the fitting process was stopped when a reasonable fit was found. The model predictions, shown as white circles in Figures 1, 2, and 3 are quite close to the observed values, despite the limited parameter search and the constraint that all parameter values excluding one (e.g., the criterion) were identical for all four groups.

Received September 16, 2004
Revision received February 23, 2005
Accepted May 4, 2005 ■