

Pairs do not suffer interference from other types of pairs or single items in associative recognition

AMY H. CRISS and RICHARD M. SHIFFRIN
Indiana University, Bloomington, Indiana

What is the source of interference on a memory test following study of a list containing different types of pairs? Many current models predict that pairs and singles of all types will jointly interfere and therefore harm memory. Such list length effects have often been observed for lists of a single-item type (e.g., a list of words). Here, we examine interference for lists containing multiple types of pairs (e.g., word–word, face–face, word–face). In three experiments, we manipulate the number of each type on the study list. In associative recognition, discrimination fell as the number of pairs of the same type rose, but the number of pairs of other types had little effect. That is, we found a list length effect within, but not between, classes of stimuli. We highlight the importance of representation and propose alternatives to current model representations that can predict such findings.

The meaning of the word *jam* in the pair *strawberry jam* clearly differs from its meaning in the pair *traffic jam*. Indeed, research has shown that memory for an item is a function of the match between the semantic context at study and that at test (Light & Carter-Sobell, 1970; Tulving & Thompson, 1973). Related to this principle is the possibility that study of word pairs, even for unrelated words, might induce configural meaning that goes beyond and may be independent of the meaning of the constituent words in isolation (see Clark & Gronlund, 1996, for a review of the independence hypothesis). Evidence for configural processing of unrelated word pairs comes from Doshier and Rosedale (1997), who found cuing advantages for triples only when all three components were studied together. Furthermore, Hockley (1992) showed that singles and pairs have different forgetting functions, and Hockley and Cristi (1996a) showed that item memory and associative memory are differentially affected by instructional manipulations.

In this article, we continue to explore configural processing by examining the effects of such processing on interference during retrieval. For example, word pairs as a class might tend to be dissimilar from single words as a class, hence reducing cross-class memory interference. Similarly, the class of word–face pairs might be dissimilar from the class of word–word pairs, and so forth. Few studies have looked at length effects that cross item–type boundaries. Gillund and Shiffrin (1981) found that the

number of studied pictures affected word recall and vice versa. However, the array of strategies used in free recall makes it difficult to come to definitive conclusions concerning the source of interference effects.

Hockley and Cristi (1996b) had participants study single items and/or pairs that were repeated various numbers of times and in various combinations. In different experiments, a single item could be repeated as both a single and as part of a pair, only as part of a pair, or as part of several different pairs. In general, participants were able to judge the frequency of single items and of pairs. Critically, they were able to make separate judgments of the frequency of an item studied alone and the same item studied in a pair. Despite participants' ability to make fairly independent judgments of frequency, more traditional memory tasks may show interference. That is, singles and pairs stored in memory may be retrieved during a traditional memory task even if participants are able to focus on a subset when instructed to do so.

In the present experiments, we gathered additional evidence regarding whether the retrieval of associations is affected by the number of single items on the study list and whether the retrieval of pairs or items from one class is affected by the number of pairs or items from another class. Specifically, we used a modified list length manipulation to measure interference between and within different classes of pair types for both single-item recognition (SR) and associative recognition (AR). In both SR and AR tasks, participants studied a list of pairs (denoted as AB, CD, EF, etc.). In SR, they were tested with a sequence of single items (A, B, X, Y, etc.), judging whether each had been studied (a target, such as A) or not (a foil, such as X). In AR, they were tested with a sequence of pairs and judged whether each had been studied as an intact pair (e.g., AB), to which they should have responded "old," or a rearranged pair composed of two items studied in different pairs (e.g., CF), to which they should have

This research was supported by NIMH Grant 12717 to R.M.S. and a National Science Foundation Graduate Fellowship to A.H.C. We thank Mark Steyvers and Ben Murdock for helpful comments on an earlier version of this manuscript. Correspondence should be directed to A. H. Criss, who is now at the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15217 (e-mail: acriss@andrew.cmu.edu), or to R. M. Shiffrin, Department of Psychology, Indiana University, 1101 E. 10th St., Bloomington, IN 47405 (e-mail: shiffrin@indiana.edu).

responded “new.” Unlike in SR, all single items in AR had been studied, so single-item familiarity could not provide a basis for correct judgments. Instead, participants had to make judgments about the relationship between the two words. Thus, this task is considered a relatively pure measure of memory for associations (Humphreys, 1976, 1978).

Survey of Global Matching Models

We now turn to a brief survey of global matching models (GMM), which have concrete representational assumptions about single items and pairs (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Humphreys, Bain, & Pike, 1989; McClelland & Chappell, 1998; Metcalfe-Eich, 1985; Murdock, 1982, 1997; Shiffrin & Steyvers, 1997). These models incorporate the common assumption that memory traces (composite or separate) consist of a vector (or matrix) of values, equivalent to a point in a high-dimensional space. We refer to the value stored in each position as a *feature*, which is equivalent to its value on some dimension. A feature is defined as a particular position of a memory probe or trace that can be aligned with a corresponding position in another memory probe or trace, in order to allow a comparison of the values in corresponding positions. Positions that align refer to the same feature, and any that do not align refer to different features.

Extant GMMs have used two basic approaches to representing single items and pairs. In one, each single item is represented as a vector of feature values, and a pair is represented as a concatenation of the single-item vectors into a double-long vector (e.g., Diller, Nobel, & Shiffrin, 2001; Hintzman, 1988; Shiffrin & Steyvers, 1997, 1998). In these *concatenation models*, SR involves matching a test item against each of the vectors (or each half of the double-long vectors if pairs were studied) and combining the matching scores into a familiarity value that is used to make a decision. All matches of traces other than the target add variability to the decision statistic, harming performance. Hence these models predict length effects, which are defined as the drop in performance as the number of nontarget traces rises. Because pairs are represented as concatenations of single-item vectors, an increase in the number of either single items or the number of pairs should reduce performance for SR.

There are two primary ways to carry out AR in the context of concatenation models. In one approach, the double-long test probe is compared with each stored double-long trace. The matching scores are again combined into a familiarity measure that is used to make a decision. An intact test pair tends to match all $2N$ features of one trace, whereas a rearranged test pair tends to match two different traces in N features each. Because familiarity is calculated as a product of evidence from each feature (in Shiffrin & Steyvers, 1997, or the cube of the evidence in Hintzman, 1988), $2N$ matching features in one trace tend to contribute more to familiarity than N matching features in each of two traces, producing above-

chance AR performance. Note, however, that for foil probes, the N matching features in each of two traces tends to contribute much more to familiarity than do the accidentally matching features in all the traces of pairs that do not contain either of the items in the test probe. This reduces the dependence of performance upon the number of these other traces, largely eliminating the prediction of length effects. To reiterate, for concatenation models the errors in AR tend to be confusions caused by the two traces of the half-matching rearranged pairs, but not confusions with the traces of the other studied pairs.

The other approach to AR in concatenation models uses a cued recall process. In the extant models, a single item is used as the recall cue. Each single member of the test pair is used as a probe cue in an attempt to recall the trace containing that item (Diller et al., 2001) or to produce a composite vector dominated by feature values of the single item that had been paired with the test item (Hintzman, 1988). Both methods involve a step in which the cue item is matched to single-item subvectors in the stored traces, and hence both predict length effects. AR performance should decrease as the number of studied pairs increases (because the number of studied single items increases in step with the number of studied pairs). It would be possible to imagine cued recall models in which the probe consists of the two test items taken together. Such a model would be similar to the joint probe strategy mentioned above. Thus, when a foil is tested, the traces in memory that would dominate retrieval would be the two half-matching traces, reducing or eliminating length effects.

A second class of models, *composite models*, represents single items and pairs as vectors, but the vector representing a pair is independent of the vector representing a single item. Thus one could describe such models as having “emergent” associative features. However, the standard versions of these models assume that the vectors representing single items and pairs are superimposed into a single summed composite memory vector (e.g., the TODAM model of Murdock, 1982, or the CHARM model of Metcalfe-Eich, 1985). These models have the interesting property of dissociating a pair from the single items of which it is composed. Yet because associative and single-item traces are stored in the same vector positions, the match of the test probe to the stored composite vector involves matching the test probe to all traces of all types. This statement holds whether the matching is direct (as, for example, in the TODAM recognition model) or due to a recall process (as, for example, in CHARM or the cued recall model in TODAM). That is, regardless of recognition or recall retrieval processes, these composite storage representations predict that increases in the number of studied single items, number of studied pairs, or both will harm performance (i.e., the list length effect). This conclusion applies to both SR and AR tasks.

The above discussion of TODAM assumes that the memory vector is zeroed (i.e., empty) at the beginning of

a study list. However, more recent versions of the model implement the continuous memory assumption: the idea that the composite memory vector contains all preexperimental experiences, as well as the current list (Murdock, 1997; Murdock & Kahana, 1993). To implement this idea, the memory vector is not zeroed prior to the experiment. Under this assumption and the assumption that context is not used during AR, Murdock (1997) showed no forgetting for pairs due to the use of context drift as the primary cause of forgetting. Our discussion assumes the original formulation of TODAM (Murdock, 1982; Weber, 1988), where list length effects are caused by the increase in variance as items are added to the memory vector along with forgetting. To foreshadow, we will find that performance for AR is *not* a function of the entire list length, but depends on the total number of pairs of the same type. We point out that TODAM can either predict no list length effect (i.e., no forgetting) for pairs (i.e., Murdock 1997) or a list length effect dependent on the total list length (i.e., Murdock, 1982). However, it should be clear that both cannot be simultaneously predicted.

Such models provide the background for the present studies. We briefly reiterate that we use single-item and associative recognition to explore the existence or absence of length effects within and between classes of pairs. The study lists contain different numbers of pairs of different classes: word–word pairs (WW), face–face pairs (FF), and word–face pairs (WF). Memory is tested using both AR and SR. Although we are primarily interested in changes in discrimination, we also report hits and false alarms. We use d_a (Macmillan and Creelman, 1991) as our measure of discrimination, but several alternative measures resulted in the same patterns of data for all experiments.¹ For SR, we ask whether memory is determined by the number of the different pair types or the number of single items. Similarly, for AR we ask whether performance is determined by the number of pairs of the same or different type (e.g., whether word–face judgments are affected by the number of word–word pairs studied) or the total number of studied pairs. Both concatenation and composite models predict that the effect of list length is determined by the total number of studied pairs for both SR and AR.

EXPERIMENTS 1 AND 2

In the following two experiments, participants studied lists of WW, WF, and FF pairs. All study lists contained the same number of single words, the same number of single faces, and the same total number of pairs. What varied across lists was the relative number of each type of pair (i.e., WF, WW, and FF). Because the total list length was held constant, both concatenation and composite models predicted no change in performance for AR or SR. The lists of pairs were studied under the same incidental instructions in both experiments, which differed only at test: Experiment 1 used AR, and Experiment 2 used SR.

General Method

Materials

Black-and-white photographs of faces were selected primarily from college yearbooks and from the Olivetti Research Database of Faces (AT&T, Cambridge, 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) background.

A set of 210 hard-to-image, low-frequency words ($M = 6.46$; Kučera & Francis, 1967) were selected, excluding any words that might be used to describe a face, a person, or a characteristic of either.

Presentation of stimuli and recording of participant data were executed on IBM-compatible personal computers using Macromedia's Authorware 5 Attain software.

Procedure

The words and faces were combined to form a study list of 120 nonoverlapping pairs of three types: word–face and face–word pairs denoted WF, word–word pairs denoted WW, and face–face pairs denoted FF. The pairing was random, such that any face could be paired with another randomly selected face or a randomly selected word for each participant. The number of each type of pair presented during the study session was varied between groups. Group A studied 40 of each pair type, Group B studied 60 WF, 30 WW, and 30 FF pairs, and Group C studied 80 WF, 20 WW, and 20 FF pairs. All types of pairs were intermixed and presented in a random order during study and test.

The participants were not informed that a memory test would follow. During each study trial, the members of the study pair were presented side by side on the monitor for 3 sec. The participants judged the degree of association between the two items using a 5-point scale. The study session began and ended with six buffer trials, two of each type.

Immediately following, the participants were given an unexpected memory test (AR in Experiment 1 and SR in Experiment 2). The participants made judgments using a 6-point confidence scale where the first 3 points corresponded to *new* and the last 3 corresponded to *old*. The test session began with six trials using the buffer stimuli, and these trials are not included in any of the reported analyses.

Experiment 1

Associative Recognition Testing of Studied Pairs

Method

Participants. The number of participants in each group was varied in order to keep the total number of observations per condition approximately equal. Of the 198 Indiana University undergraduates who participated for either course credit or \$6.00, there were 38 in Group A, 66 in Group B, and 94 in Group C.

Procedure. The participants received an unexpected AR memory test. Test pairs were presented one above the other (in contrast to the study phase, in which the pairs had been presented one beside the other). All test pairs consisted of one item that had been studied on the left and one that had been studied on the right, but the test position (top vs. bottom) was not correlated with the study position (left vs. right). The test contained an equal number of intact and rearranged trials of each pair type. Rearranged test pairs were constructed within pair type (i.e., each face in a rearranged FF pair contained faces that were studied in two separate FF pairs; WW and WF foils were composed by the same method). Both members of a study pair contributed to different rearranged pairs. The number of study pairs limited the number of possible test pairs, and consequently the associative recognition test consisted of 40 pairs of each type for Group A, 28 pairs of each type for Group B, and 20 pairs of each type for Group C.

Results

For all analyses, we used an alpha level of .05. We were primarily interested in changes in discrimination; although we did not have predictions concerning changes that might occur for hits and false alarms considered separately, we analyzed this data. Hit rates (HR) did not differ between groups for WW [$F(2,195) = .50$, $MS_e = .03$], WF [$F(2,195) = .56$, $MS_e = .03$], or FF pairs [$F(2,195) = 1.93$, $MS_e = .06$], nor did false alarm rates (FAR) for WW [$F(2,195) = .44$, $MS_e = .03$], FF [$F(2,195) = 2.06$, $MS_e = .03$], or WF pairs [$F(2,195) = 2.45$, $MS_e = .03$], as shown in Table 1.

An analysis of d_a shows a within-pair type category length effect. That is, discrimination for WF pairs decreased with an increase in the number of WF pairs [$F(2,191) = 4.01$, $MS_e = .49$]. Likewise, discrimination of FF pairs decreased with an increase in the number of FF pairs [$F(2,191) = 4.02$, $MS_e = .46$]. The effect for WW pairs failed to reach significance [$F(2,194) = 1.44$, $MS_e = .55$]. The smaller length effects for WW pairs is curious but consistent with the finding of little forgetting for pairs over a relatively short time period in the continuous recognition paradigm (Hockley, 1992) and the very small or absent list length effects found for SR by Dennis and Humphreys (2001). For longer study–test delays (i.e., as short as 30 min), Hockley and Consoli (1999) have shown equal retention levels for item and associative information. Perhaps longer study–test delays would help maximize the length effect for WW pairs. The three groups of bars in Figure 1 give results for FF, WW, and WF pairs in that order. Within each group, the bars are in descending order of the number of pairs of that type on the study list. The bars generally increase in height from left to right, indicating a category length effect within pair type. The separate analysis of hits and false alarms did not show systematic and significant changes with category length and could have been due to changes in criterion placement between lists (among other factors). This study does not allow us to attribute the source of the discrimination changes to hits or false alarms.

In the present design, the number of test trials was not held constant between groups, theoretically allowing learning during testing to differentially affect the groups.

Table 1
Associative Recognition Hit Rates and False Alarm Rates for Experiment 1

Type of Pair	Number of Pairs	HR	FAR
FF	40	.49	.31
	30	.46	.24
	20	.53	.25
WW	40	.67	.23
	30	.70	.22
	20	.70	.24
WF	80	.69	.23
	60	.71	.18
	40	.73	.18

Note—W, word; F, face.

Although the participants surely encoded something in memory during each test trial, it seems likely that these traces are weak and have little impact on the present data. For example, strengthening items via repetition does not harm recognition or cued recall performance for other items from the list (i.e., the null list strength effect; Ratcliff, Clark, & Shiffrin, 1990), suggesting that additional intact test pairs would not harm performance much. Of greatest importance, this confound does not predict the data. Performance is best for WF pairs in the condition with the most test pairs (i.e., Group A with 40 study pairs of each type). Nevertheless, we carried out additional analyses to ease concerns about this issue. The above statistical analyses were recomputed while restricting the data to the first 60 test trials of each condition. The qualitative patterns of data are identical. Again, none of the HRs or FARs changed with the number of pairs on the lists [although the WF FAR is marginally significant, $F(2,194) = 2.93$, $MS_e = .03$, $p = .053$; all other F s < 2.16 and p s $> .12$]. For d_a , we again find no effect for WW pairs ($F < 1$) but a decrease in performance as the number of studied pairs of the same type increase for FF pairs [$F(2,192) = 6.26$, $MS_e = .54$] and WF pairs [$F(2,182) = 2.69$, $MS_e = .49$, $p = .07$, although marginally significant].² Taking into account all of these issues and noting that we replicate these findings in Experiment 3, we believe the confound to be unfortunate but immaterial to the main thesis of this article: that performance in AR is determined by pairs of the same type and not by pairs of a different type or by single items.

Although discussion is deferred until the presentation of Experiment 2, it should be noted that the existence of within-type length effects implies that interference in associative recognition is not simply determined by the total number of pairs on the list (as predicted by the models), because the total number of pairs was held constant.

Experiment 2

Single-Item Recognition Testing of Studied Pairs

Method

Participants. One hundred twenty-five Indiana University undergraduates received course credit or \$6.00 for participating in a 35-min session. Groups A and B each had 42 participants, and Group C had 41.

Procedure. The participants received an unexpected single-item recognition test consisting of 120 single faces and single words randomly intermixed. These consisted of 20 studied items from each pair type, 30 word foils, and 30 face foils. Of the 20 test items from WF pairs, half were faces and half were words.

Results

HRs and FARs for Experiment 2 are shown in Table 2. HRs for words were slightly greater than for faces [$F(1,122) = 5.54$, $MS_e = .03$] and for items studied in WF pairs [$F(1,122) = 13.44$, $MS_e = 1.41$]. FARs were higher for faces than for words [$F(1,122) = 104.06$, $MS_e = .02$]. Importantly, there was no evidence for a change in $P(\text{old})$ as the number of the different pair types varied

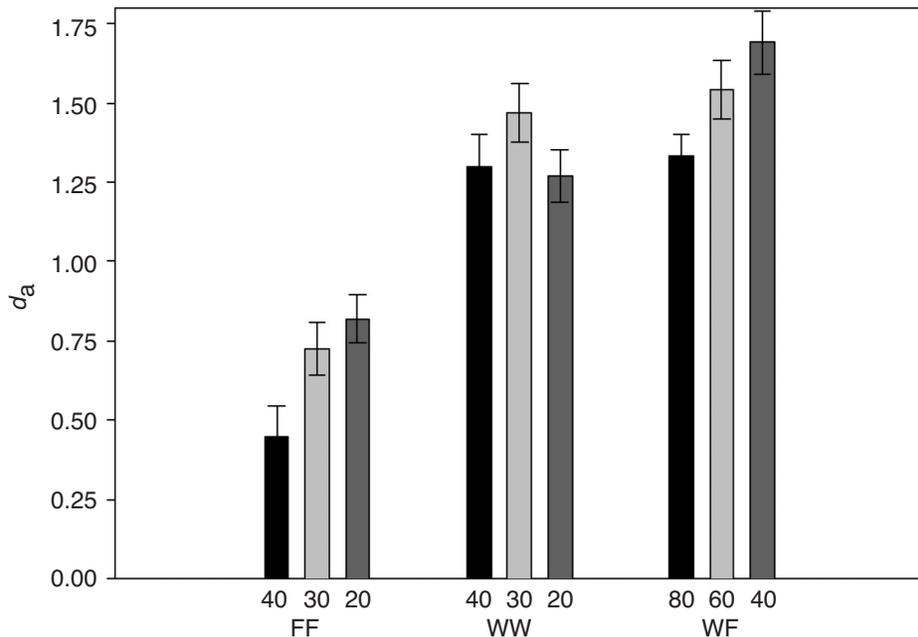


Figure 1. Discrimination as a function of pair type (FF, WW, or WF) and number of studied pairs of the same type in Experiment 1. The number under each bar indicates the number of studied pairs of that type. Error bars in all graphs are standard errors. F, face; W, word.

for the HR [$F(2,122) = .33$, $MS_e = .06$] or the FAR [$F(2,122) = .96$, $MS_e = .04$]. The value of d_a , graphed in Figure 2, was higher for words than for faces [$F(1,97) = 68.15$, $MS_e = .76$] but did not change as a function of the relative proportion of each type of studied pair [$F(2,97) = 1.23$, $MS_e = 1.15$]. This result, different from the pattern of performance for AR in Experiment 1, is predicted by extant concatenation and composite models. For these models, the variation in the relative number of different types of pairs across groups would not have affected single-item (or AR) performance, given that the total number of single faces and single words remained constant.

Discussion of Experiments 1 and 2

In AR but not SR, performance for a given pair type improves as the relative proportion of pairs of that type

decreases. This result by itself of course implies a differentiation by type—all pairs are not equal in their interfering effects. If similarity of the pair types to each other were constant, within-type length effects would not be present. Alternative explanations based on differential study can be ruled out because the test type was postcued. Because the study conditions were identical in the two experiments (regardless of the later type of testing), it would be hard to argue that different study strategies were responsible for the patterns observed. The different patterns of performance, then, must be a result of some difference between AR and SR.

A great number of studies demonstrate list-length effects in recognition and recall tasks using lists of a single type of item (but see Dennis & Humphreys, 2001). A number of studies have also shown category length effects in free recall when categories are mixed in lists (e.g., Tulving & Pearlstone, 1966). The length effects mentioned thus far have involved discrimination changes. A different type of category length effect has been observed in a few recognition studies using several categories per list. These effects involved a parallel change in both hits and false alarms without a change in discriminability. Such effects were found for several categories of words defined by semantic or visual/pronunciation similarity (Criss & Shiffrin, 2004a; Shiffrin, Huber, & Marinelli, 1995; Sommers & Lewis, 1999). The present results are stronger because we find discrimination-based length effects within pair type, even when the total length of the list and the total number of pairs, faces, and words was held constant for each list.

Table 2
Single-Item Recognition Hit Rates and False Alarm Rates for Experiment 2

Group	HR	FAR
Faces		
A (40WF, 40WW, 40FF)	.72	.36
B (60WF, 30WW, 30FF)	.71	.34
C (80WF, 20WW, 20FF)	.68	.31
Words		
A (40WF, 40WW, 40FF)	.73	.19
B (60WF, 30WW, 30FF)	.73	.16
C (80WF, 20WW, 20FF)	.73	.16

Note—W, word; F, face.

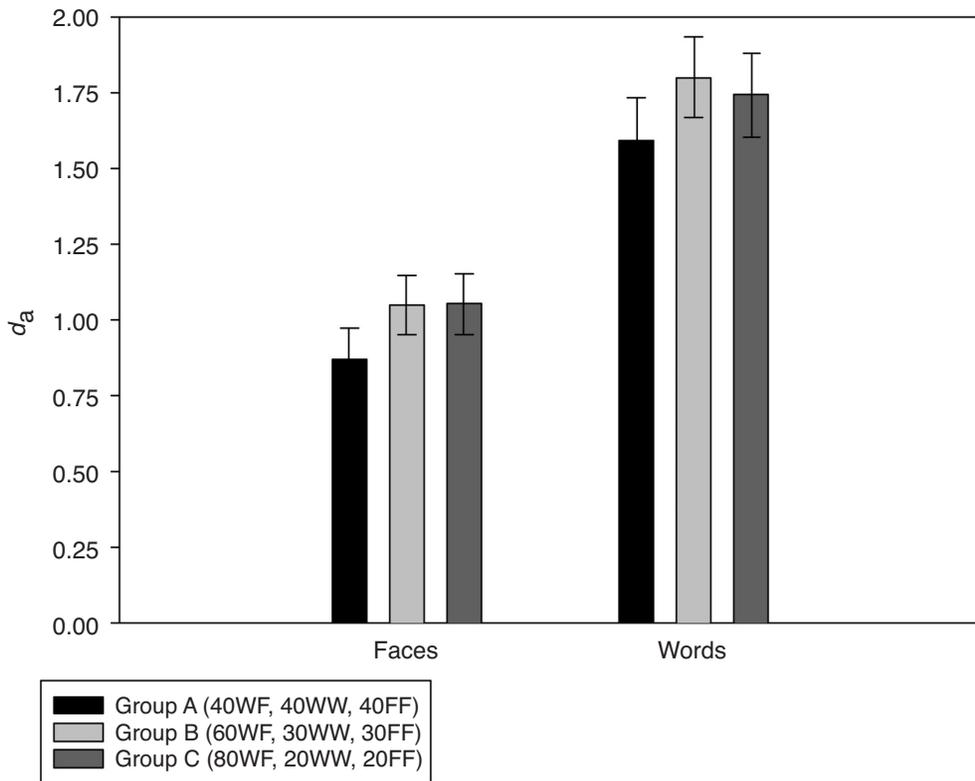


Figure 2. Discrimination for single-item recognition in Experiment 2 as a function of the item type and the experimental group. Note that Group A studied 40 of each pair type, Group B studied 60WF, 30WW, and 30FF, and Group C studied 80WF, 20WW, and 20FF. F, face; W, word.

To model their findings, Shiffrin et al. (1995) and Criss and Shiffrin (2004a) suggested that the increase in familiarity and $P(\text{old})$ is caused by the similarity between stored traces of items in the category of the test item and the test item itself. Familiarity and $P(\text{old})$ increased as the number of such traces increased. By design, the majority of studied items were unrelated to any one test item, and each of these also contributes variance but less than traces of items in the test-item category. The accumulated noise due to traces of items outside the test-item category produced most of the variance. As a result and as demonstrated with each of their models, category size produced a change in the $P(\text{old})$ but no change in discrimination. In the present studies, the similarity between the three pair types and the two item types was likely much lower (e.g., the pair types are likely dissimilar to each other), and the relative number of other-category items was lower. Thus, it is plausible that the relatively few, dissimilar other-category pairs contributed less variance than did the similar same-category pairs.

Finally, we have indirect evidence concerning the similarity of pairs composed of words and faces (in any of the three possible combinations) to the single items composing those pairs. Pairs of items of a type different from the test pair type nonetheless contain the same type of single items (e.g., both FF and WF pairs contain faces),

but the number of such pairs does not have an effect. That is, the number of single items in those other-type pairs does not reduce performance for AR. Furthermore, performance levels for pairs are not predictive of performance levels for the single items composing them. For example, consider those participants in Group A who studied an equal number of each pair type. Inspection of the rightmost bar of Figure 1 and the black bars in Figure 2 reveal that this group of participants had the best performance for associative recognition of WF pairs (i.e., Figure 1) but the worst performance for recognition of single Ws and Fs (i.e., Figure 2). This suggests a possible tradeoff between the encoding of item and associative information (Hockley & Cristi, 1996a; McGee, 1980; Murdock, 1982) and indirectly supports the hypothesis of separate representations for pairs and single items.

Due to the implications of the findings of the first two studies on representation and their potential to constrain theories of associative recognition, we replicate and further explore them in the next experiment.

EXPERIMENT 3

Experiment 1 covaried the number of each pair type between participants with the result that AR performance changed with the number of pairs of the same type. The

present study was designed to test this length effect within participants rather than between participants. Of more substantive interest, this experiment was designed to discover whether and how adding items of one type to the study list affects performance for a constant number of studied pairs of another type. In the previous studies, adding items of one type required removing items of the other type, in order to maintain a constant list length. In this experiment, each participant completed three study–test blocks. For each participant, the number of pairs of one type was held constant across blocks, and the number of another type of pair (and consequently, the total list length) was varied. On the basis of the previous experiment, we expect that adding pairs of the same type would harm performance. In contrast, both classes of models predicted that performance would be determined by the total number of pairs.

Method

Participants. A total of 325 Indiana University undergraduates participated in return for either course credit or \$6.00 for the 30-min session. Each student participated in only one of the six conditions described next.

Materials. Materials were drawn from those used in Experiments 1 and 2.

Procedure. There were six between-participants conditions, varying in the type of pairs used. Each participant had three study–test blocks, described shortly. Each group received just two pair types: (WF, FF), (WF, WW), (WW, WF), (WW, FF), (FF, WF), (FF, WW), where the first pair type in each set indicated the type that always had 20 pairs, and the second pair type in each set indicated the pair type that had either 0, 10, or 20 pairs across lists for that participant. We refer to the pair type that had 20 members on each study list as the *constant pair type* and the pair type that had 0, 10, or 20 pairs on the list as the *varied pair type*. An example of one temporal order of events for the first of these six groups (WF, FF) is shown in Figure 3. The figure depicts a particular order of the three list types (denoted A, B, C), but the order was randomly chosen for each participant. Where there is a comma separating the pair types within a row, the two types were actually presented in a randomly mixed order.

Each pair was studied for 3 sec, during which time the participants judged the degree of association between the two items. Two buffer trials began and ended each study list (not shown in the example), and the order of lists and the stimuli on each list were randomly chosen for each participant. The length of puzzle activity separating the study and test phases varied in order to maintain a constant study–test lag for the critical test items (those above the horizontal line in the figure). Specifically, the total time between the first study item and the first test item was constant for each study–test block.

Note that each participant received one *pure list* (containing only one pair type, as in List A in Figure 3) and two *mixed lists* (containing two pair types, as in Lists B and C in Figure 3). To equate the mixed lists for the average amount of switching between pair types and the average study–test lag, we constructed the mixed lists with the constraint that the first 20 pairs of each of these study lists (i.e., those above the horizontal line in Figure 3) included 10 pairs of each type. The test pairs were constructed from these study pairs, as described next.

All test trials were AR, constructed as in Experiment 1. The critical tests consisted of 10 intact and 10 rearranged pairs taken from the first 20 pairs of each study list. In the cases where there were two types of pairs, half of the test items were from each type. To keep participants from noticing that items from certain study positions were never tested, one intact and one rearranged test were included from the second block of 10 study items in List B, and two intact and two rearranged tests were included from the second study block of 20 items in List C (these tests were not analyzed). All test pairs were randomly intermixed.

Results and Discussion

The same basic pattern of data was obtained for the six between-participants groups (which varied only in the two pair types used). In the varied conditions, there was an interaction between category length and participant group on HR [$F(5, 319) = 3.19, MS_e = .04$] and a marginally significant interaction for d_a [$F(5, 194) = 2.21, MS_e = .43, p < .10$] both due to a larger effect of length for FF pairs. There were no other significant interactions. The different pair types had different overall levels of performance, but the main focus in this experiment is on the

<u>List A</u>	<u>List B</u>	<u>List C</u>
20WF	10WF, 10FF	10WF, 10FF
Puzzle activity	10WF	10WF, 10FF
Puzzle activity	Puzzle activity	Puzzle activity
AR test	AR test	AR test

Figure 3. The basic design for Experiment 3 is depicted. In this example, WF is the constant pair type, and FF is the varied pair type. List A is a pure list, and Lists B and C are mixed. The comma indicates that those sets of pairs are randomly intermixed during study. Test pairs are constructed from the study pairs above the horizontal line. F, face; W, word.

pattern of results across conditions. We therefore present the data collapsed across groups: HRs and FARs are shown in Table 3, and d_a is shown in Figure 4.³

The varied pair type is a slightly modified replication of Experiment 1. If we held the number of pairs of another type constant, adding pairs of the same type should have harmed discrimination. As expected, we replicated the within-pair type length effect. As shown in the right two bars of Figure 4, discrimination was higher when the lists had 10 (light bar) pairs than when the list had 20 (dark bar) pairs [$F(1,194) = 6.16, MS_e = .43$] of the varied type. The change in discrimination is primarily due to a decrease in the HR as the number of pairs of the same type increased [$F(1,319) = 24.57, MS_e = .04$]. The FARs did not change [$F(1,319) = .30, MS_e = .04$].

The constant pair type is an extension of Experiment 1, where we ask: What is the effect of adding pairs of another type while holding the number of pairs of the same type constant? FARs for the constant pair types did not change across conditions [$F(2,638) = .52, MS_e = .03$], so the findings described below are manifest in d_a and the HRs. The main effect of list (0, 10, or 20 pairs of the other type) on both d_a and the HR was significant [$F(2,442) = 31.09, MS_e = .44; F(2,638) = 19.71, MS_e = .04$, respectively], as such, post hoc analyses are reported below.

There are two primary findings. First, consider the mixed lists (the lighter two bars on the left side of Figure 4). We found no difference in discrimination as additional pairs of another type were added to the study list ($F < 1, p = .70$). HRs mimic the pattern of d_a , as shown in Table 3. That is, when the number of pairs is held constant, we find no difference between adding 10 pairs or 20 pairs of another type [$F(1,319) = 1.65, MS_e = .05$].

In brief, adding pairs of the same type as the test pair harms performance, but adding pairs of another type does not. This supports a model where pairs of different types have distinct representations and are dissimilar to one another (despite sharing single items from the same class). Furthermore, the lack of interference of pairs as another type is added suggests that the number of single items does not alter AR performance (because adding pairs of another type must add single items). This result suggests that single items and pairs are represented and retrieved separately.

Comparing the black bar to the two light bars on the left side of Figure 4 demonstrates the second important

finding: an advantage for pure lists (those containing pairs of a single type) over mixed lists. Discrimination was greater for the pure list than for the list with an additional 10 pairs of another type [$F(1,258) = 59.39, MS_e = .46$] and for the list with 20 such pairs [$F(1,244) = 68.38, MS_e = .37$]. Again, this is largely due to a higher HR for items in pure lists than for items in mixed lists [for List B, $F(1,319) = 23.45, MS_e = .03$; for List C, $F(1,319) = 41.67, MS_e = .03$]. Performance drops when pairs of another type are added, relative to a pure list, but the number of such pairs (10 or 20) does not matter, as described above. By design, the amount of switching between pair types at study and at test is the same for all test items from the mixed lists, so it appears that the drop in performance is a result of switching between two pair types.

Our data cannot distinguish whether such switching costs occurred at study or test. A plausible argument can be made for an effect occurring at study. The pure list advantage could be due to different strategies, one for each pair type, for carrying out the incidental study instructions. If some encoding time is lost in switching from the strategy for one type to the encoding strategy for another type, switching between types would reduce performance compared with the case where switching does not occur. This argument places the pure list advantage at study and is similar to those found in the task switching literature (e.g., Rubinstein, Meyer, & Evans, 2001; Sohn & Anderson, 2001).

On the other hand, one could imagine that it takes time and effort to focus on the relevant subset of items in memory for each test probe. For example, it might take time to construct a cue focusing on a particular pair type, such as FF. When tests alternate between types, this process could produce a deficit compared with the case where all tests were of the same type (as for the pure list). Relevant evidence in the literature comes from studies of categorized free and cued recall (for a review, see Raaijmakers & Shiffrin, 1980) and from studies of the fan effect (e.g., J. R. Anderson & Reder, 1999; M. C. Anderson & Neely, 1996). In free recall, for example, there is a tendency to recall rapidly from one category (e.g., fruits), but then search slowly for another category to output (assuming the category names are not provided). Similar processes may be at work in studies of the fan effect. In these studies, participants learn (or preexperimentally know) several facts related to some topic, and these facts might fall into different subcategories. Such studies have shown that response times to verify facts depend on the number of subcategories and the number of facts within the relevant subcategory, but not the number of facts within the irrelevant subcategories (McCloskey & Bigler, 1980; Reder & Anderson, 1980). We leave it to future research to determine whether the switching costs observed in the present study occur at study, retrieval, or both.

GENERAL DISCUSSION

We have shown that different classes of pairs do not interfere with one another during retrieval. That is, AR

Table 3
Associative Recognition Hit Rates and False Alarm Rates
for Experiment 3

Pair Type	HR	FAR
Constant		
0 others	.73	.20
10 others	.66	.21
20 others	.64	.20
Varied		
10 others	.76	.21
20 others	.68	.20

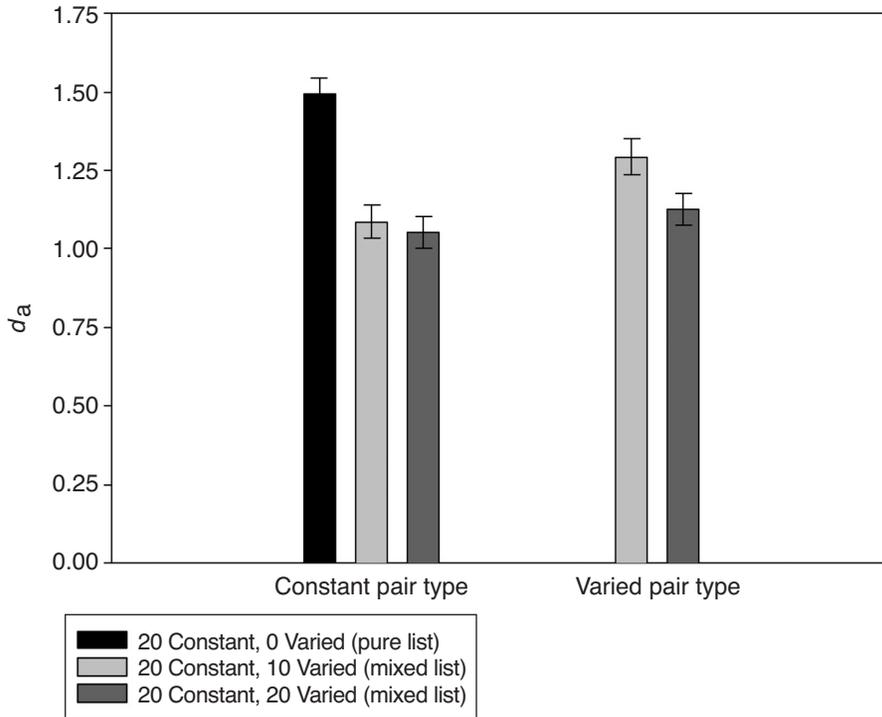


Figure 4. Discrimination as a function of study condition for Experiment 3.

performance is a function of the number of pairs of the same type and not the total list length. Furthermore, we have evidence suggesting that singles do not interfere with pairs during AR. As pairs of another type are added to the study, list performance does not change, even though these pairs contain single items that are common to the test pair. Both findings are inconsistent with current formulations of composite and concatenation models. The concatenation models are not sufficient for this data, because pairs are simply singles stored in the same memory trace and thus any manipulation changing performance for pairs must be similarly reflected in performance for singles. Composite models are also unable to account for the current data, but for a different reason. While some of these models assume different (and orthogonal) memory traces for pairs and their constituent singles, these models assume that all memory traces are summed into one memory trace, which forces interference to be a function of the total list length. Next, we discuss related data followed by a description of alternative representations that could be implemented in models in order to more fully account for the data.

The inference that single items and pairs maintain a form of functional separation is consistent with a number of previous studies. As discussed earlier, SR and AR are differentially sensitive to instructions (Hockley & Cristi, 1996a), have different forgetting rates (Hockley, 1992), and have different rates of improvement with study time (Clark & Shiffrin, 1992). Judgments of fre-

quency indicate that people are generally able to make separate judgments for pairs and singles, even when they have words in common (Hockley & Cristi, 1996b). The present results fit nicely with this research, providing additional evidence for the separation of items and associations during both storage and retrieval.

Another set of data posing problems for each of these model classes was obtained by Kelley and Wixted (2001). Participants studied pairs one or six times and were tested with intact pairs, rearranged pairs, and new–new pairs, under instructions to call only intact pairs “old.” Two findings were of particular relevance. First, the HR for strong pairs exceeded that for weak pairs, but the weak and strong rearranged FARs did not differ. Concatenation models cannot account for this without additional assumptions, because the strong rearranged test items will match their half-matching vectors more than will the weak rearranged test items. On the contrary, such a finding is consistent with a model like TODAM, in which pairs are independent of the items from which they are constructed. In this model, pair AB is no more similar to pair AD, with which it shares an item, than to another pair EF, with which it shares no item (see Weber, 1988). Thus, the strength of the relevant studied pair has no effect on the FAR. Second, the (not differing) FARs to rearranged weak and rearranged strong pairs were greater than the FAR to new–new pairs (consistent with previous findings of Humphreys, 1976, and Clark & Shiffrin, 1992). This finding is consistent with concatenation

models. When memory has two traces that half-match the rearranged test pair, FARs should be higher than when memory has no traces that match the rearranged test pair at all. But this is inconsistent with a composite model like TODAM for the reason mentioned above; namely, the test pair is equally unrelated to a memory vector that incorporates two half-matching traces and a memory vector that contains no half-matches (Weber, 1988).

Data from Kelley and Wixted (2001), the present set of experiments, and the Hockley studies mentioned earlier seem to require some degree of independence between the representations of single items and pairs. Depending on the details, such a model is likely to predict no difference in FARs to weak and strong rearranged pairs. To account for the lower FAR to new–new foils than to rearranged foils, one could simply assume that participants adopt a strategy when faced with such testing conditions. First, each item is used as a probe. If neither item matches memory, a “new” response is given. If at least one item matches memory, an associative probe is used (such as the convolution in TODAM), and a decision is made on the basis of that match. Given the nature of the test items, this strategy seems sensible. Gronlund and Ratcliff (1989) and Nobel and Shiffrin (2001) have shown that time to discriminate new–new foils is faster than to rearranged foils, perhaps lending some support to this type of model. It should be noted that Kelley and Wixted claim that their data is best explained in terms of competition between a familiarity process and a recall-to-reject process. We have explained their data in terms of a familiarity process with independent representations for singles and pairs. The growing importance of considering issues of representation in addition to the processes involved in AR is addressed further in the “Recall Processes in AR” section. Data from the present set of experiments seem to require a more generous modification, and we now turn to such extensions.

In general terms, what is needed is a way to represent varying levels of similarity even for items that seem nominally similar (i.e., that WF and WW pairs both contain words but nevertheless seem rather dissimilar). The degree of similarity is assumed to be a joint function of encoding processes and the stimuli (see Criss & Shiffrin, 2004b, for an example of how encoding tasks alter the word-frequency effect, another finding thought to be attributable to retrieval processes). For example, as Paivio (1971) argued, visual and verbal materials may have different types of memory codes (or features). However, pairs created from items within the same domain or items from different domains may also form dissimilar traces due to encoding processes. The exact mechanisms leading to similar or dissimilar memory traces is left for future development. We simply illustrate three different representations that could arise from the unspecified encoding processes.

First, suppose that each of the classes (W, F, WW, WF, FF) is represented by the same features in the same re-

gion of the memory vector, but with values that are similar within class and dissimilar between classes.⁴ Suppose, in addition, that a pair is coded as three traces, one for each separate item, and one for the pair. An example of the traces stored in memory for study of a WF pair is shown in the top panel of Figure 5. Here, features identifying the list context are stored with each trace and are denoted C. A test of any type will strongly match the traces that encode that type and weakly match the others. Note that this is similar to a representation proposed in Murdock (1982; Model 4). Our proposal simply requires that varying degrees of similarity be built into the sets of memory traces; namely, traces are similar within class but not between classes. How to best implement varying degrees of similarity is unclear. For example, the REM model (Shiffrin & Steyvers, 1997) has a natural floor to dissimilarity when features are chosen randomly. Under these circumstances, the likelihood ratio approach implemented in REM has a natural and optimal criterion for making an old/new decision. For a given item to be more similar to other items of its own type than another type, the within-class similarity must be greater than random. This means that within-class foils will have high matching values, and a higher criterion would be needed. Although plausible, this assumption loses one advantage of the likelihood ratio approach: the optimal criterion setting. Most other models also assume that items are only randomly similar, and straying from this assumption may require substantial changes in the proposed mechanisms.

A closely related representation assumes that pair features are distinct for each type of pair and different from single-item features. The idea is to use *separate regions*

[W ₁ F ₁ C]
[W ₁ C]
[F ₁ C]

[W F WW FF WF C]
[W ₁ F ₁ W ₁ F ₁ C]

[W ₁ Wcode C]
[F ₁ Fcode C]
[W ₁ F ₁ WFcode C]

Figure 5. An example of each of the three alternate representations described in the text. Each panel shows the stored memory traces following study of a WF pair. Features identifying the list context are denoted “C.” The top panel illustrates the model with high within-type similarity and low between-type similarity. The middle panel illustrates the separate regions model, with the general representation shown in the first row followed by an example for study of a WF pair in the second row. The bottom panel illustrates the type code model. W, word; F, face.

of feature values in the representation for different item types. For the present application, there would be six regions: one for context, C, and one region for each of W, F, WW, FF, and WF, as shown in the first row of the middle panel in Figure 5.⁵ For storage of a WF pair, as shown in the second row of the middle panel in Figure 5, the regions in which feature values are stored would be C, W, F, and WF. It is convenient to elaborate the vector to be a matrix, so multiple items of the same type can be stored in the same event (e.g., two Ws would need to be stored for a WW pair). This extension can be thought of as an extension of Model 2 proposed in Murdock (1982), where memory was assumed to consist of two composite vectors: one for singles and one for pairs. It is possible to implement this type of representation in both classes of models. One benefit of this model is that it allows a single-memory trace (which would take the form of a matrix in REM, for example) to represent a complex event, whereas the other models break one event into separate traces. Although these first two representations are similar in many respects, in the separate feature model just described, the concept of feature (and feature value) is more strongly reified because different item types are assigned different features. One can imagine techniques used to derive actual features for encoding faces and words (e.g., Griffiths & Steyvers, 2003; Landauer & Dumais, 1997; Steyvers, 2001), and it seems plausible that these will differ in kind.

Finally, consider a representation that makes use of *type codes*: Suppose that a set of common features is used to encode all item types and that these have equal between- and within-type similarity. However, in addition to these features and those encoding list context, suppose that there is a region of features used to encode the type (i.e., whether the stored item is WW, WF, FF, etc.). An example of the traces stored for a WF study pair is shown in the bottom panel of Figure 5. This representation is straightforward in concatenation models. In order to have just one type code per trace, it is simplest to assume three separate traces for each studied pair: one for each item of the pair and one for the pair itself. At retrieval, list context and type code are first used as probes to activate a subset of list traces for that type of item (followed by matching the item features as usual). The type cue would tend to activate only the traces of type and length effects that would be restricted to the test type. For composite models, the use of this representation could be implemented in a four-way convolution (in TODAM, or a four-way matrix multiplication in the matrix model) including each item, the list context, and the type code. Note that the use of a type code could apply to any number of attributes of a study item, such as the gender of the voice producing the study word. Underwood (1969) discussed a similar concept ("class attributes" in his words). In support of this notion, he used the example that when searching for a technical term, one does not generate the name of a colleague. In some sense, explicitly using type codes (or class attributes) is one

step toward defining context and separating it into its component parts.

Any of the representations discussed above, if implemented in extant models, would have the property of both separating pairs from the singles from which they were constructed and likewise separating various subclasses from each other. As such, these representations are able to handle the data discussed earlier showing the separation of items and pairs during both storage and retrieval. The Kelley and Wixted (2001) study requires the additional assumption that a strategic use of item information may be evoked when test items include new–new foils. In this strategy, each item is first compared with memory. Pairs are rejected if both single items are below some familiarity criterion. If either single item exceeds this criterion, memory is probed with the associative code and an intact–rearranged discrimination is made.

Recall Processes in AR

So far, we have assumed for simplicity that AR involves decisions based on a recovered familiarity value (cf. Dyne, Humphreys, Bain, & Pike, 1990). However, there may be reasons to think that AR is carried out with a recall-like process. If so, could the proposed representational approach fit the present results? The answer depends on the cue used to probe memory. If the test pair is coded in terms of associative features and these are used to probe memory (i.e., for a WW test pair, only the WW features are used to probe memory), sampling would tend to be restricted to items of the same type and produce the correct length or null-length effects. In other words, if an associative cue is used, regardless of whether a familiarity or recall process or some combination is used, length effects will depend on the similarity between the cue and traces of the same type (assuming any one of the above representations is adopted). On the other hand, some researchers have proposed recall models in which the memory probe is one of the individual items composing the pair, akin to cued recall (e.g., Rotello & Heit, 2000; Rotello, Macmillan, & Van Tassel, 2000). In this case, sampling would be based on all traces that contain the single-item feature in the probe. For example, a WW test would involve probing with a single W, activating all traces containing W features. Thus, length effects would clearly depend on the number of WW and WF traces, which is not in accord with the present findings. In conclusion, our data do not permit a clear choice between a recall-based and a familiarity-based model of AR. Clearly, any successful model will need to incorporate some form of the representational assumptions we have proposed.

Several studies have suggested that recall processes are involved in SR and/or AR.⁶ For example, the Nobel and Shiffrin (2001) data exhibited a much closer match between RT distributions for AR and for cued recall than between RT distributions for AR and for SR. These findings were interpreted to imply that both AR and cued recall involve an extended search process producing slow

retrieval. This is, however, suggestive rather than definitive, because the slow time course of retrieval in AR could be due to the time required to generate the associative encoding that is used to probe memory. In fact, this generation-time argument could be used to explain slow AR response times, even when a familiarity model is assumed (see Gronlund & Ratcliff, 1989, for a similar proposal). Thus, the Nobel and Shiffrin results do not provide definitive evidence that could be used to assess the representational issues we have discussed.

One other source of relevant evidence concerning AR comes from studies of forced-choice AR. Clark, Hori, and Callan (1993) presented pairs (AB, CD, EF, GH, IJ, etc.) for study and gave a 3-AFC test. The OLAP condition contained choices that shared a studied item such as AB, AD, or AF. The NOLAP conditions contained choices not sharing any items such as AB, CF, or GJ. If, as our data suggest, AB coding is unrelated to A coding, AB coding might be unrelated to AC coding. If so, our approach would predict no difference between NOLAP and OLAP. Clark et al. (1993) found a NOLAP advantage and argued that this was due to the use of cued recall; the NOLAP case provides more single items to use as cues. However, Clark and Hori (1995) found similar performance for NOLAP and OLAP for longer study lists and suggested that participants may have abandoned the single-item probe strategy. It may be that our designs were similar in that they prevented participants from using a single-item probe recall strategy.

In summary, the representational approach we have suggested is consistent with the extant literature even if AR is carried out by a recall process, as long as the recall probe is composed of configural pair information. This discussion highlights the importance of considering both representation and process. Many recent articles have drawn conclusions regarding the processes underlying AR (e.g., see Macho, 2004, for a review of several recent examples) without much regard for representation. These proposals are incomplete without equal consideration of the underlying representation. Here, we have outlined three different representations that could be adopted in any extant model to form a more complete model of AR.

Summary

Interference, measured by list length effects, was found within each class of pairs (WW, FF, WF) but not across classes. In addition, switching between pair types harmed performance. The results were taken to imply separate representations for these various types of items and pairs. Several methods were discussed by which different representations could be achieved and implemented in several GMMs.

REFERENCES

- ANDERSON, J. R., & REDER, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186-197.
- ANDERSON, M. C., & NEELY, J. H. (Eds.) (1996). *Interference and inhibition in memory retrieval*. San Diego: Academic Press.
- AT&T LABORATORIES CAMBRIDGE (1994). *The Olivetti research database of faces* [Electronic database]. Available from www.uk.research.att.com/facedatabase/html.
- 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. (1995). List length and overlap effects in forced-choice associative recognition. *Memory & Cognition*, *23*, 456-461.
- CLARK, S. E., HORI, A., & CALLAN, D. E. (1993). Forced-choice associative recognition: Implications for global-memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 871-881.
- CLARK, S. E., & SHIFFRIN, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition*, *20*, 580-598.
- CRISS, A., & SHIFFRIN, R. M. (2004a). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, *111*, 800-807.
- CRISS, A., & SHIFFRIN, R. M. (2004b). Interactions between study task, study time, and the low-frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 778-786.
- DENNIS, S., & HUMPHREYS, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.
- DILLER, D. E., NOBEL, P. A., & SHIFFRIN, R. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 414-435.
- 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, & Cognition*, *16*, 813-824.
- GILLUND, G., & SHIFFRIN, R. M. (1981). Free recall of complex pictures and abstract words. *Journal of Verbal Learning & Verbal Behavior*, *20*, 575-592.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *19*, 1-67.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. Oxford: Wiley.
- GRIFFITHS, T. L., & STEYVERS, M. (2003). Prediction and semantic association. In S. Becker, S. Thrun, & K. O. Ghahramani (Eds.), *Advances in neural information processing systems* (pp. 11-18). Cambridge, MA: MIT Press.
- GRONLUND, S. D., & RATCLIFF, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, 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.
- HOCKLEY, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, & 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.
- HUMPHREYS, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition*, *4*, 221-232.
- HUMPHREYS, M. S. (1978). Item and relational information: A case for context-independent retrieval. *Journal of Verbal Learning & Verbal Behavior*, *17*, 175-187.

- HUMPHREYS, M. S., BAIN, J. D., & PIKE, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, **96**, 208-233.
- KELLEY, R., & WIXTED, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 701-722.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **102**, 211-240.
- LIGHT, L. L., & CARTER-SOBELL, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning & Verbal Behavior*, **9**, 1-11.
- MACHO, S. (2004). Modeling associative recognition: A comparison of two-high-threshold, two-high-threshold signal detection, and mixture distribution models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 83-97.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MALMBERG, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 380-387.
- MCCLELLAND, J. L., & CHAPPELL, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, **105**, 724-760.
- 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.
- 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, & Cognition*, **19**, 689-697.
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.
- NOBEL, P. A., & SHIFFRIN, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 384-413.
- PAIVIO, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- POLLACK, I., & NORMAN, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, **1**, 125-126.
- RAADMAKERS, J. G., & SHIFFRIN, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207-262). New York: Academic Press.
- RATCLIFF, R., CLARK, S. E., & SHIFFRIN, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, & 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.
- 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., & VAN TASSEL, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language*, **43**, 67-88.
- RUBINSTEIN, J. S., MEYER, D. E., & EVANS, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 763-797.
- SHIFFRIN, R. M., HUBER, D. E., & MARINELLI, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 267-287.
- 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: Oxford University Press.
- SOHN, M.-H., & ANDERSON, J. R. (2001). Task preparation and task repetition: Two-component model of task switching. *Journal of Experimental Psychology: General*, **130**, 762-778.
- SOMMERS, M. S., & LEWIS, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory & Language*, **40**, 83-108.
- SMALL, S. A., NAVA, A. S., PERERA, G. M., DELAPAZ, R., MAYEUX, R., & STERN, Y. (2001). Circuit mechanisms underlying memory encoding and retrieval in the long axis of the hippocampal formation. *Nature Neuroscience*, **4**, 442-449.
- STEVVERS, M. (2001). Modeling semantic and orthographic similarity effects on memory for individual words (Doctoral dissertation, Indiana University). *Dissertation Abstracts International*, **61**, 11B.
- TULVING, E., & PEARLSTONE, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, **5**, 381-391.
- 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.
- WEBER, E. U. (1988). Expectation and variance of item resemblance distributions in a convolution-correlation model of distributed memory. *Journal of Mathematical Psychology*, **32**, 1-43.
- YONELINAS, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, **25**, 747-763.
- YONELINAS, A. P., KROLL, N. E. A., DOBBINS, I. G., & SOLTANI, M. (1999). Recognition memory of faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review*, **6**, 654-661.
- ZEINEH, M. M., ENGEL, S. A., THOMPSON, P. M., & BOOKHEIMER, S. Y. (2003). Dynamics of the hippocampus during encoding and retrieval of face-name pairs. *Science*, **299**, 577-580.

NOTES

1. Our measure of discrimination, d_a (Macmillan & Creelman, 1991) requires use of the slope of the zROC calculated for each condition for each participant. Due to a small number of observations per condition and idiosyncratic use of the confidence scale, the slope is sometimes undefined. In such cases, d_a cannot be calculated and that participant is eliminated from the analysis for the relevant comparisons. To be sure that the patterns of data presented in this article were not due to violations of the assumptions required by d_a or to the elimination of those participants with undefined slopes for at least one of the conditions being compared, we analyzed the data for each experiment using the following methods: d' (Green & Swets, 1966), G (Nelson, 1984), A' (Pollack & Norman, 1964), and $HR - FAR$. All the measures we used resulted in the same pattern of discrimination (although tests of significance sometimes varied), confirming the robustness of the findings. We will provide these analyses upon request.
2. Restricting the data to a subset of test trials resulted in elimination of additional participants due to elimination of participants with undefined slopes (see note 1). This loss of power is likely why these statistical tests are less reliable despite showing the same qualitative pattern.
3. A complete table of values is available upon request.
4. An unpublished study had participants study single faces and words. The results showed no cross-class interference. That is, adding words to a list harmed performance (i.e., d_a) for words, not for faces.

Likewise, study of additional faces harmed performance for faces, but not for words. This led us to a model where faces and words were stored separately.

5. It is intriguing to note that fMRI studies have shown different areas of the hippocampus to be active during encoding of faces, names, and face–name pairings (Small et al., 2001) and different areas to be active during the encoding and retrieval stages of learning face–name pairs (Zeineh, Engel, Thompson, & Bookheimer, 2003). However, topographic separation does not necessarily imply functional independence.

6. Another main line of support for the use of recall in AR comes from studies of the shape of the ROC. Yonelinas (1997; Yonelinas, Kroll, Dobbins, & Soltani, 1999) has obtained linear rather than curvi-

linear ROCs in AR tasks and used such a finding to argue for the use of a recall process (but see Kelley & Wixted, 2001, for evidence of curvilinear ROCs in AR). We used our confidence rating data to produce ROC curves and observed curvilinear ROCs in all conditions. We do not present these findings in this article because they are not informative. Curvilinear ROCs are not as diagnostic as linear ones, because they could arise due to the presence of a wide variety of noise and guessing processes, even if the underlying retrieval process was based on recall (Malmberg, 2002).

(Manuscript received October 24, 2003;
revision accepted for publication March 16, 2004.)