THE REPRESENTATION OF SINGLE ITEMS AND ASSOCIATIONS IN EPISODIC MEMORY

Amy H. Criss

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Departments of Psychology and Cognitive Science

Indiana University

July 2004

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

July 2004

Richard M. Shiffrin, Ph.D.

Chairperson

Robert M. Nosofsky, Ph.D.

William K. Estes, Ph.D.

Joseph E. Steinmetz, Ph.D.

Acknowledgments

As I write this last piece of my dissertation, I can't help but feel slightly disappointed. Apparently upon completion of a PhD, it is necessary to flee the nest and move onto to new challenges. But leaving this department and this lab in particular is not an action I take with enthusiasm. Our lab has been a source of constant intellectual stimulation, boundless support, and just plain fun. I cannot imagine a better environment in which to grow as a researcher and as a person. Thus, I am somewhat saddened that I have finally earned the right to leave. I feel incredibly fortunate to have been a part of this lab and I am eternally grateful to the many post-docs, graduate students, and visitors for having shared their friendship and unabashed honest opinions about my research and the field in general. My appreciation for my advisor Rich Shiffrin cannot be overstated and I thank you for everything.

Our lab dynamics are similar to the department as a whole and I'd like to thank the entire department, and members of my committee in particular, for their constant willingness to answer questions and their unrelenting ability to challenge me to be a better researcher.

I grew up in a family that always believed in me but were nevertheless exceedingly proud of even the smallest of achievements. They gave me the confidence and motivation to continue learning even in the face of the negative reviews, unclear data, and comments such as "Amy, this is a complete disaster." I cannot thank you enough.

Finally, to Jayson for not taking me seriously, for learning to listen to lab talk even at Nick's, for forcing me to have fun when I just wanted to work, and for the immeasurable way in which you love me, I thank you.

Preface

This manuscript is the culmination of several years of research aimed at understanding how associations are represented in memory, how that representation relates to the representation for single items, and how to best implement these principles within the framework of existing mathematical models. The manuscript is written in such a way that each of the three parts could be read (mostly) independent of the others. The first part has recently been accepted for publication in Memory & Cognition. It is the initial set of studies using various types of pairs (word-word, word-face, and face-face) as a tool for measuring interference and thus drawing conclusions about the similarity between different classes of pairs. In this set of studies, we found that performance in associative recognition is determined by the number of within-class pairs and not affected by pairs of another class. Single item recognition, however, was determined by the total number of singles and not affected by pair-type. We explain why no existing model can account for this pattern of data and describe modifications of existing models that would be able to capture the pattern of data. Part II seeks to gather converging evidence for this important finding using a different paradigm. The primary finding is that single items repeated during study have no affect on associative recognition performance when those items were studied in a different pair-type. Only when single items are repeated in the same pair-type do we see interference, in the form of a tendency to call any pair old if it contains repeated items. Again, single item recognition shows a different pattern of results. Namely, single items that were studied in an earlier study list show equal amounts of interference regardless of whether the repetitions occurred in the same or different type of pair. The model described in Part I captures this pattern data both qualitatively and quantitatively. Finally, in Part III we manipulate instructions at study in order to better understand the degree to which these findings depend on the stimulus type and/or encoding strategies. We find that both the independence between single item and pair representations and the dissimilarity of the three types of pairs depend on encoding strategies. Though this set of studies is not entirely conclusive, it seems that conditions encouraging participants to relate the study pairs in a unique way led to the patterns described above. Whereas, when Ss are not given specific instructions on how to study the items, we see neither of these independencies.

Abstract

What is the representation of pairs in memory? Some models make the assumption that a pair is represented as the two component items bound together in time. Others assume emergent pair information beyond the information contained in each singleton. In fact, the present set of studies will show that neither of these assumptions will suffice. Using an associative recognition task requiring discrimination between two items studied together and two items studied as members of different pairs, we found that discrimination fell as the number of studied pairs of the same type rose, but the number of studied pairs of other types had little effect. That is, we find a list length effect within but not between classes of stimuli, when we define the classes of stimuli to be word-face, word-word, and face-face pairs. On the contrary, single item recognition is not influenced by pair-type. The model types employing the assumptions described above were unable to account for this pattern of data, leading to the development of a novel set of models. A test of the new model was carried out in another empirical setting using two successive study lists with repetitions of some items and some pairs across lists. This design required that intact pairs in the recent list be distinguished from rearranged pairs as well as intact pairs from the previous list. The results showed that between-list confusions only occurred for pairs of the same type, even when the constituent single items were repeated across lists, confirming the previous conclusions and model. On the other hand, we find the same pattern of confusions for single item recognition, regardless of the type of pair(s) in which the single items were studied. The model developed to handle the first paradigm proved capable of predicting the new results in quantitative

detail. Finally, we briefly address the degree to which encoding strategies determine the independence of singles and pairs and the independence of these pair-types.

Table of Contents

Part I

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

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 Dosher and Rosedale (1997) who found cuing advantages for triples only when all three components were studied together. Further, Hockley (1992) showed that singles and pairs have different forgetting functions and Hockley & Cristi (1996a) showed that item and associative memory are differentially affected by instructional manipulations.

In this paper, 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 & 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 singles that were studied as singles and singles studied as members of a pair. Despite Ss 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 Ss are able to focus in on a subset when instructed to do so.

The current experiments gather 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 use a modified list length manipulation to measure interference between and within different classes of pair-types for both single item (SR) and associative recognition (AR). In both our SR and AR tasks, participants study a list of pairs (denoted as AB, CD, EF, etc.). In SR they are 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 are tested with a sequence of pairs and judge whether each had been studied as an intact pair (e.g., AB), to which they should respond "old," or a rearranged pair composed of two items studied in different pairs (e.g., CF), to which they should respond "new." Unlike SR, all single items in AR have been studied, so single item familiarity cannot provide a basis for correct judgments. Instead, participants must 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; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1999; Murdock, 1982; 1997; Humphreys, Bain, & Pike, 1989; Hintzman, 1988; Metcalfe-Eich, 1985). 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 to be 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 the values in corresponding positions to be compared. 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 by a concatenation of the single item vectors into a double long vector (e.g. Hintzman, 1988; Shiffrin & Steyvers, 1997; 1998; Diller, Nobel, & Shiffrin, 2001). 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 for a decision. All matches of traces other than the target add variability to the decision statistic, reducing performance. Hence these models predict length effects, defined as the drop in performance as the number of non-target 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 to 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 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, reducing (or eliminating) any dependence on list length.

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, Nobel, & Shiffrin, 2001) or 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 sub-vectors 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 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 composite models have the interesting property of dissociating a pair from the single items of which it is comprised. 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 reduce 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 pre-experimental 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 additional items are added to the memory vector in addition to forgetting. To foreshadow, we will find that performance for AR is <u>not</u> a function of the entire list length, but depends on the total number of pairs of the same type. We have pointed 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 item types. The study lists contain different numbers of items of different classes: word-word pairs (WW), face-face pairs (FF), and word-face pairs (WF). Memory is tested using both AR and SR. Though 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 though note that several alternative measures resulted in the same patterns of data for all experiments.[1] For SR, we ask whether memory is determined by the numbers of the different pair-types or the number of single items. Similarly, for AR we ask whether performance is determined by the numbers 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 contain the same number of single words, the same number of single faces, and the same total number of pairs. What varies across lists is the relative number of each type of pair (i.e., WF, WW, and FF). Because the total list length is held constant, both concatenation and composite models predict no change in performance for AR or SR. The lists of pairs were studied under the same incidental instructions in both experiments. The experiments differed only at test: Experiment 1 used AR, and Experiment 2 used SR.

General Methods

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; Kucera & 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 was 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 non-overlapping 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.

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 s. 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, participants were given an unexpected memory test (AR in Experiment 1 and SR in Experiment 2). 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

Methods

Participants

The number of participants in each group 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.

Test 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 analysis we used an alpha level of .05. We are 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 ($F(2, 195)=.50$, $MSE=.03$, $p=.622$), WF ($F(2, 195)=.48$, $MSE=.03$, $p=.621$), or FF pairs ($F(2, 195)=2.37$, $MSE=.06$, $p=.096$), nor did false alarm rates (FAR) for WW ($F(2, 195)=.57$, $MSE=.03$, $p=.569$), FF ($F(2, 195)=2.178$, $MSE=.03$, $p=.116$). FARs did vary between groups for WF pairs ($F(2, 195)=3.123$, $MSE=.03$, $p=.046$), 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 of the number of WF pairs ($F(2, 191)=4.01$, $MSE=.49$, $p=.02$). Likewise, discrimination of FF pairs decreased with an increase in the number of FF pairs ($F(2, 191)=4.02$, $MSE=.46$, $p=.02$). The effect for WW pairs failed to reach significance ($F(2, 194)=1.44$, $MSE=.55$, $p=.264$). 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 occasionally found for words in SR (Dennis & Humphreys, 2001). For longer study-test delays (i.e., as short as 30 minutes), Hockley & 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. That the separate analysis of hits and false alarms did not show systematic and significant changes with category length was not unexpected, 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. Though participants surely encoded something in memory during each test trial, it seems likely that these traces are weak and have little impact on the current data. For example, strengthening items via repetitions 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 analysis to ease concerns about this issue. The above statistical analyses were re-computed 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 (though the WF FAR is marginally significant, $\underline{F}(2, 194)=2.93$, $\underline{MSE}=.03$, $\underline{p}=.053$; all other $\underline{F}$'s<2.16 and $\underline{p}$'s$>.12$). For d_a, we again find no effect for WW pairs ($\underline{F}<1$), but a decrease in performance as the number of studied pairs of the same type increase for FF pairs ($\underline{F}(2, 192)=6.26$, $\underline{MSE}=.54$, $\underline{p}=.002$) and WF pairs ($\underline{F}(2, 182)=2.69$, $\underline{MSE}=.49$, $\underline{p}=.07$, though 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 manuscript: 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

Methods

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.

Test Procedure

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 faces ($F(1, 122)=5.54$, $MSE=.03$, $p=.02$) and for items studied in WF pairs ($F(1, 122)=13.44$, $MSE=1.41$, $p<.001$). FARs were higher for faces than words ($F(1, 122)=104.06$, $MSE=.02$, $p<.001$). Importantly, there was no evidence for a change in P(old) as the number of the different pair-types varied for the HR ($F(2, 122)=.33$, MSE=.06, p=.723) or the FAR (F(2, 122)=.96, MSE=.04, p=.384). The value of d_a, graphed in Figure 2, was higher for words than faces (F(1, 97)=68.15, MSE=.76, $p<.001$), but did not change significantly as the number of each type of pair varied between groups (F(2, 97)=1.23, MSE=1.15, p=.298). 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 post-cued. Because the study conditions were identical in the two studies (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.

To model their findings, Shiffrin et al. (1995) and Criss & Shiffrin (2004a) suggested that the increase in familiarity and P(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(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 their model, category size produced a change in the P(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 items contributed less variance (if any at all) compared to the similar same-category items resulting in a discrimination based change in performance as a function of the number of same-category pairs.

Finally, we have indirect evidence concerning the similarity of pairs comprised of words and faces (in any of the three possible combinations) to the single items comprising 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. Further, performance levels for pairs are not predictive of performance levels for the single items comprising them. For example, consider those participants in Group A that studied an equal number of each pair-type. Inspection of the right-most 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 trade-off 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 co-varied 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 if 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 completes three study-test blocks. For each participant, the number of pairs of one type is held constant across blocks and the number of another type of pair (and consequently, the total list length) is varied. Based on the previous experiment, we expect that adding pairs of the same type will harm performance for that pair-type and not affect performance for other types of pairs. In contrast, both classes of models predict performance will be determined by the total number of pairs.

Methods

Participants

A total of 325 Indiana University undergraduates participated in return for either course credit or $6.00 for the 30 min. session. Each participant participated in only 1 of the 6 conditions described next.

Materials

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

Procedure

There were six between-participant 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 indicates the type that always had 20 pairs, and the second pair-type in each set indicates 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, 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 s 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 receives 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 ten 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 ten study items in List B, and two intact and two rearranged tests were included from the second study block of twenty items in List C (these tests were not analyzed). All tests pairs were randomly intermixed.

Results & Discussion

The same basic pattern of data was obtained for the six between-participant groups (that varied only in the 2 pair-types used). In the varied conditions, there was an interaction between category length and participant group on HR ($\underline{F}(5, 319)=3.19$, $\underline{MSE}=.04$, $\underline{p}=.008$) and a marginally significant interaction for d_a ($F(5, 194)=2.21$, $\underline{MSE}=.43$, $\underline{p}=.055$) 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 pattern of results across conditions. We therefore present the data collapsed across group: 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. Holding the number of pairs of another type constant, adding pairs of the same type should harm 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 pairs (dark bar), $\underline{F}(1, 194)=6.16$, $\underline{MSE}=.43$, $\underline{p}=.014$ 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 ($\underline{F}(1, 319)=24.57$, $\underline{MSE}=.04$, $\underline{p}<.001$). The FARs did not change ($\underline{F}(1, 319)=.30$, $\underline{MSE}=.04$, $\underline{p}=.587$).

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, MSE=.03, p=.594 so the findings described below are manifest in d_a and the hit rates. The main effect of list (0, 10, or 20 pairs of the other type) on both d_a and the hit rate was significant (F(2, 442)=31.09, MSE=.44, p<.001; F(2, 638)=19.71, MSE=.04, p<.001, respectively), as such, post hoc analysis 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 items of the other type were added to the study list (t(238)=.472, p=.637). 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 of another type and 20 pairs of another type (t(324)=1.373, p=.171).

In brief, adding pairs of the same type as the test pair harms performance but additional 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). Further, 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 for the pure list was greater than the list with an additional 10 pairs of another type ($t(236)=7.78$, $p<.001$) and the list with 20 such pairs ($t(249)=8.15$, $p<.001$). Again, this is largely due to a higher HR for those items in pure lists compared to mixed lists (for list B, $t(324)=4.803$, $p<.001$; for list C $t(324)=6.518$, $p<.001$). 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, then 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 when 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 (see Raaijmakers & Shiffrin, 1980 for a review) and from studies of the fan effect (e.g., Anderson & Reder, 1999; 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 pre-experimentally know) several facts related to some topic and these facts might fall into different sub-categories. Such studies have shown that response times to verify facts depend on the number of sub-categories and the number of facts within the relevant sub-category, but not the number of items within the irrelevant sub-categories (Reder & Anderson, 1980; McCloskey & Bigler, 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 performance is a function of the number of pairs of the same type and not the total list length. Further, we have evidence suggesting that singles do not interfere with pairs during associative recognition. 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. This assumption 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, item and associative recognition 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 frequency indicate that people are generally able to make separate judgments for pairs and singles, even when they share words in common (Hockley & Cristi, 1996b). The present results fit nicely with this research, providing additional evidence for the separation of item and associative information during both storage and retrieval.

Another set of data posing problems for each of these model classes was obtained by Kelley & Wixted (2001). Participants studied pairs 1 or 6 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 the HR 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 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 not consistent with a composite model like TODAM for the same 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 & Wixted (2001), the current 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 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, then 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 & Shiffrin (2001) have shown that time to discriminate new-new foils is faster than rearranged foils, perhaps lending some support to this type of model. It should be noted that Kelley & Wixted claim 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 single and pairs. The growing importance of considering issues of representation in addition to the processes involved in AR is addressed further in the section titled "Recall Processes in AR." Data from the current 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 to be rather dissimilar). The degree of similarity is assumed to be a joint function of encoding processes and the stimulus materials (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 region 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 varying degrees of similarity be built into the sets of memory traces. Namely, traces are similar within class but not between. 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 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 would be 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 without much problem. 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. Though 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 used to encode faces and words (e.g., Steyvers, 2001; Griffiths & Steyvers, 2003; Landauer & Dumais, 1997) 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 these have equal between- and within-type similarity. However, in addition to these features and those encoding list context, suppose 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 that type and length effects would be restricted to the test type. For composite models, the use of this representation could be implemented in a 4-way convolution (in TODAM, or a 4-way matrix multiplication in the Matrix model) including each item, the list context, and the type-code. Note 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) discusses 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 sub-classes 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 & 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 to 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 this is correct, 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), then 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 cues of the same type (assuming any one of the above representations is adopted). On the other hand, many have proposed recall models in which the memory probe is one of the individual items comprising the pair, akin to cued recall (e.g., Rotello & Heit, 2000; Rotello, Macmillan, & Van Tassel, 2000; Shiffrin & Steyvers, 1998). 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, not in accord with the present findings. In conclusion, our data do not permit a clear choice between a recall and a familiarity-based model of AR. What is clear is that 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 reaction time distributions for AR and cued recall than between AR and SR. These findings were interpreted to imply that both AR and cued recall involved 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 (2001) results do not provide definitive evidence that could be used to assess these issues.

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, AF. The NOLAP conditions contained choices not sharing any items such as AB, CF, and GJ. If, as our present data suggest, AB coding is unrelated to A coding, then AB coding might be unrelated to AC coding. If so our present 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 the 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 comprised 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 the 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 items (F, W, 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.

Footnotes

[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 paper were not due to violations of the assumptions required by d_a or to the elimination of those participants with undefined slopes for one at least one of the conditions being compared, we analyzed the data for each experiment using the following methods: d-prime (Green & Swets, 1966), G (Nelson, 1984), A' (Pollack & Norman, 1964), and H-FA. All the measures we used resulted in the same pattern of discrimination (though 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 Footnote 1). This loss of power is the likely reason 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, while adding words to a list harmed performance (i.e., d_a) for words, additional words did not harm performance for faces. Likewise, study of additional faces harmed performance for faces but not for words. This lead 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 are active during encoding of faces, names, and face-name pairings (Small et al., 2001) and different areas are 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 curvilinear 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).

Table 1.

Associative Recognition Hit and False Alarm Rates for Experiment 1.

Number of Study Pairs	HR	FAR
	FF	
40	.49 (.04)	.31 (.03)
30	.46 (.03)	.24 (.02)
20	.53 (.02)	.25 (.02)
	WW	
40	.67 (.03)	.23 (.02)
30	.70 (.02)	.22 (.02)
20	.70 (.02)	.24 (.02)
	WF	
80	.69 (.02)	.23 (.02)
60	.71 (.02)	.18 (.02)
40	.73 (.02)	.18 (.02)

Table 2.

Single Item Recognition Hit and False Alarm Rates for Experiment 2.

	HR	FAR
Faces		
Group A (40WF, 40WW, 40FF)	.72 (.02)	.36 (.03)
Group B (60WF, 30WW, 30FF)	.71 (.02)	.34 (.03)
Group C (80WF, 20WW, 20FF)	.68 (.02)	.31 (.03)
Words		
Group A (40WF, 40WW, 40FF)	.73 (.02)	.19 (.02)
Group B (60WF, 30WW, 30FF)	.73 (.02)	.16 (.02)
Group C (80WF, 20WW, 20FF)	.73 (.02)	.16 (.02)

Table 3.

Associative Recognition Hit and False Alarm Rates for Experiment 3.

	HR	FAR
Constant Pair Type		
0 Others	.73 (.01)	.20 (.01)
10 Others	.66 (.01)	.21 (.01)
20 Others	.64 (.02)	.20 (.01)
Varied Pair Type		
10	.76 (.01)	.21 (.01)
20	.68 (.01)	.20 (.01)

Figure Captions

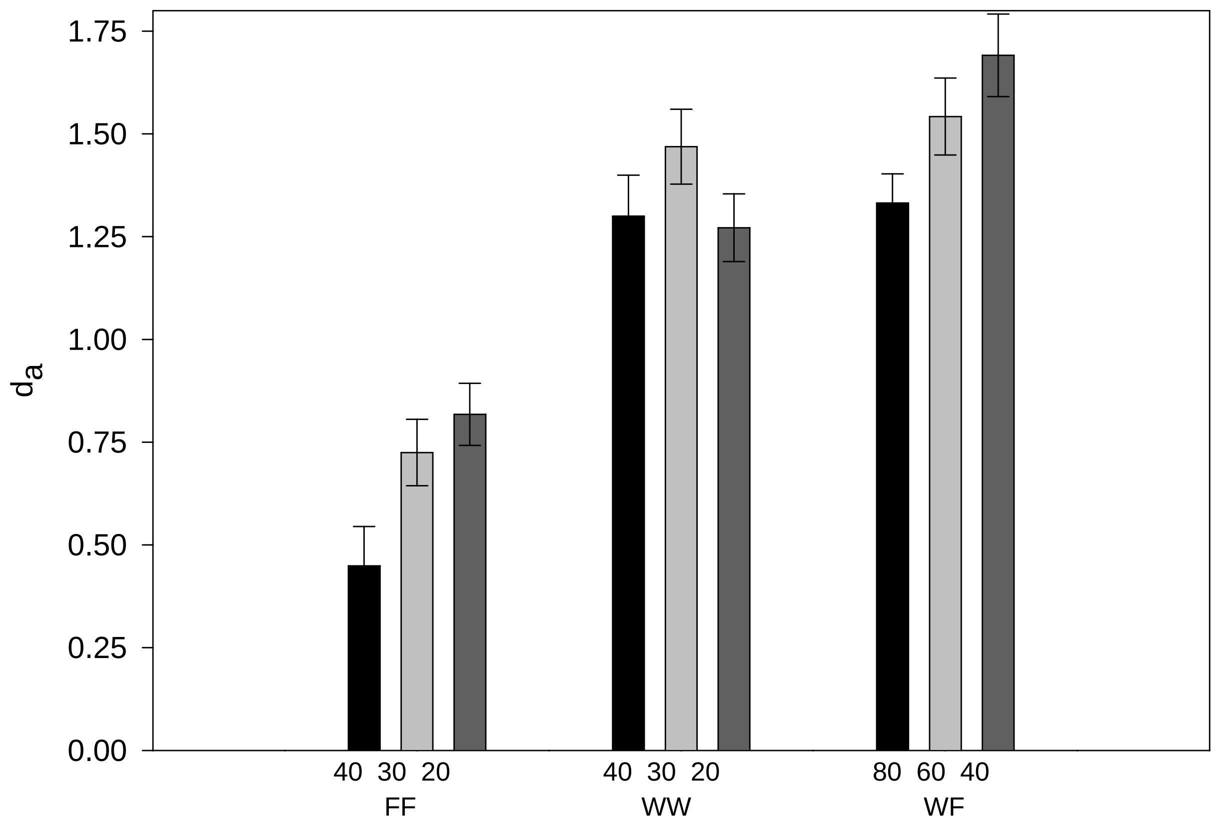
Figure 1. Discrimination in associative recognition as a function of pair-type (FF, WW, or WF) and number of studied pairs of the same type in Experiment 1. The number under each bar indicates the number of studied pairs of that type. Error bars in all graphs represent one standard errors above and one below the mean.

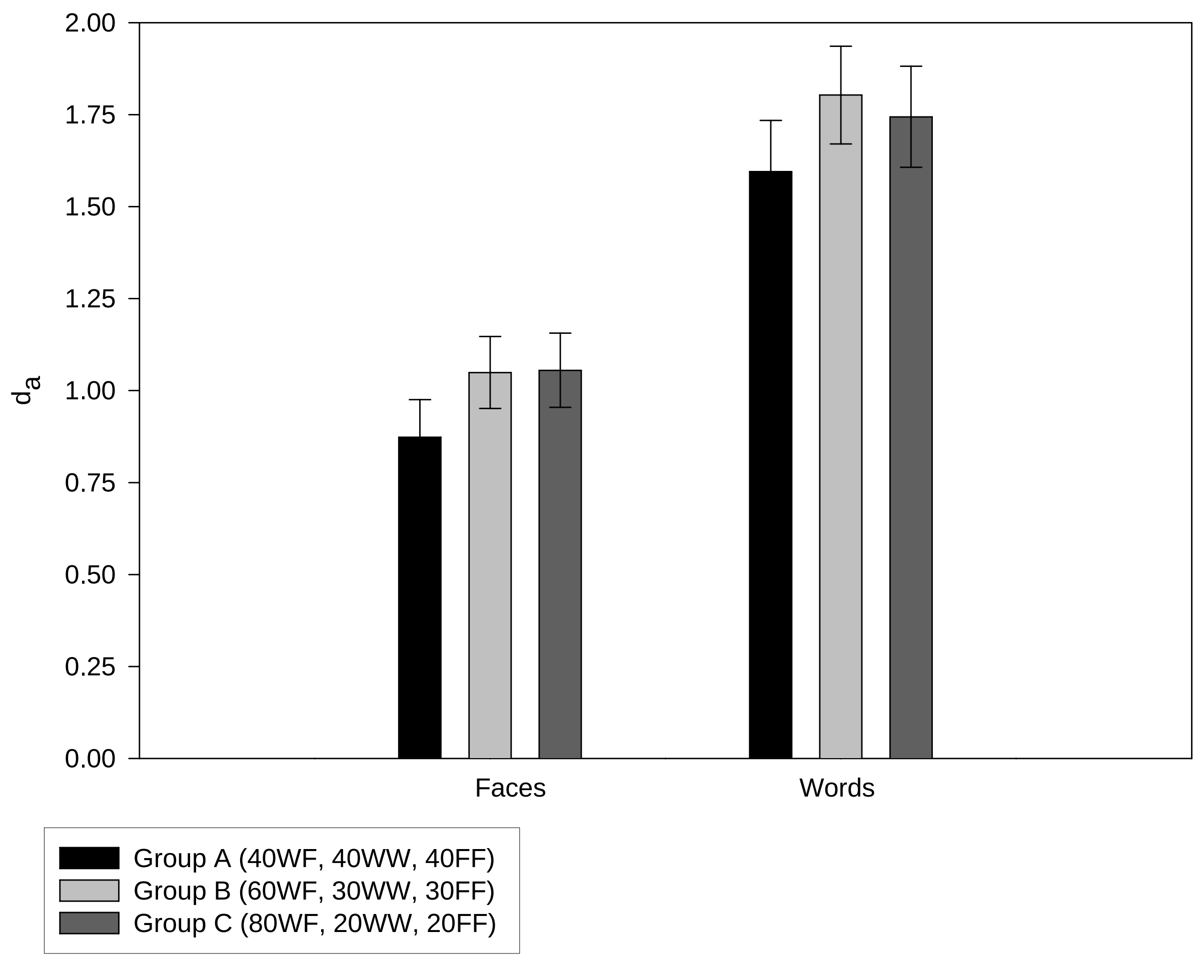
Figure 2. Discrimination for single item recognition of Experiment 2 as a function of the item type and the experimental group. Note that Group A studied 40 of each pair-type, Group B studied 60WF, 30FF, and 30WW and Group C studied 80WF, 20FF, and 20WW.

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 those study pairs above the horizontal line. See the text for further explanation.

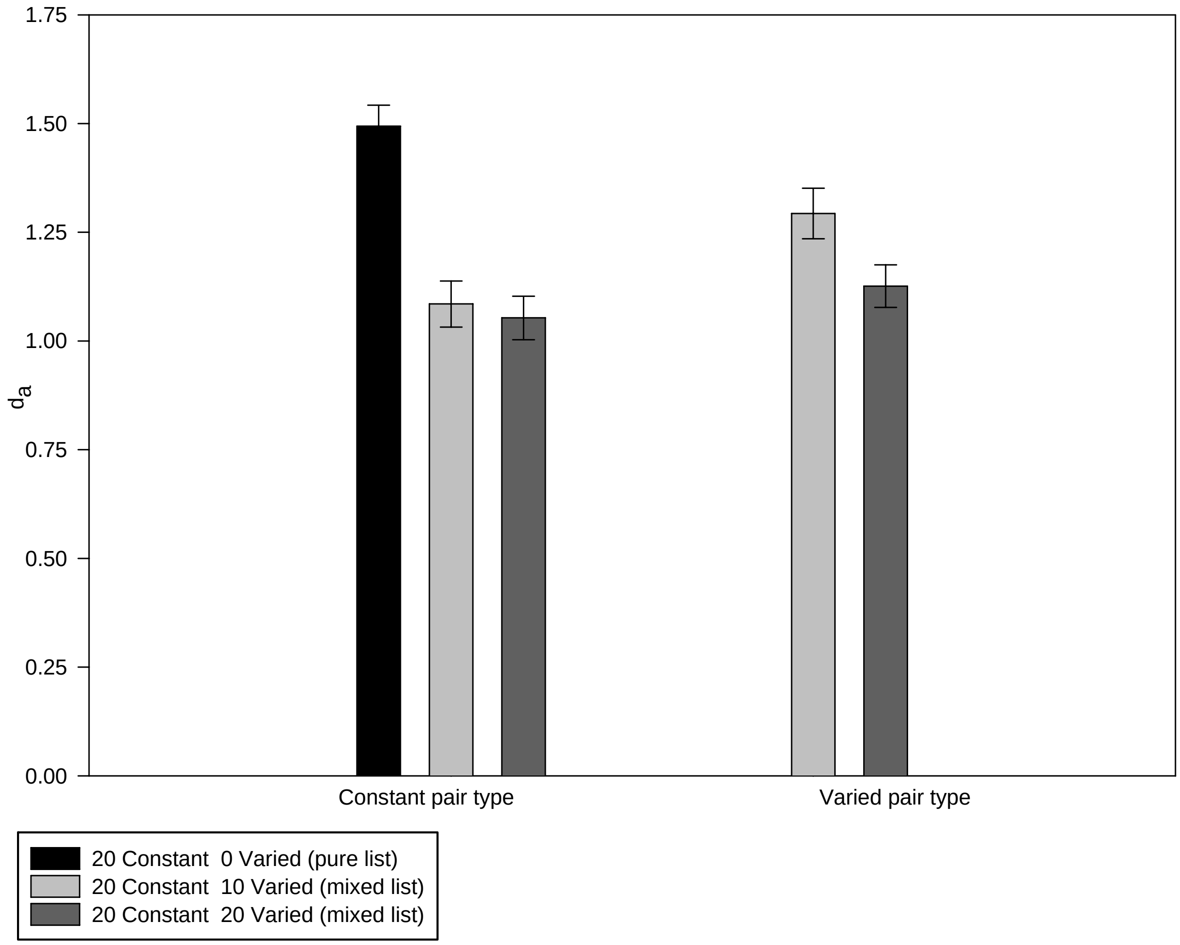
Figure 4. Discrimination as a function of study condition for Experiment 3.

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. The set of features identifying the list context is 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. The general representation is 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.


1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
d a
40 30 20
FF
40 30 20
WW
80 60 40
WF


2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
d_a
Faces
Words
Group A (40WF, 40WW, 40FF)
Group B (60WF, 30WW, 30FF)
Group C (80WF, 20WW, 20FF)

List A	List B	List C
20 WF	10 WF, 10 FF	10 WF, 10 FF
Puzzle activity	10 WF	10 WF, 10 FF
Puzzle activity	Puzzle activity	Puzzle activity
AR test	AR test	AR test



1.75
1.50
1.25
1.00
d_a
0.75
0.50
0.25
0.00
Constant pair type
Varied pair type
20 Constant 0 Varied (pure list)
20 Constant 10 Varied (mixed list)
20 Constant 20 Varied (mixed list)

$[\ W_1F_1 \quad C\ ]$

$[\ W_1 \quad C\ ]$

$[\ F_1 \quad C\ ]$

$[\ W \quad F \quad WW \quad FF \quad WF \quad C\ ]$

$[\ W_1 \quad F_1 \quad \quad \quad W_1F_1 \quad C\ ]$

$[\ W_1 \quad Wcode \quad C\ ]$

$[\ F_1 \quad Fcode \quad C\ ]$

$[\ W_1F_1 \quad WFcode \quad C\ ]$

Part II

Three-phase Recognition for Single Items and Associations

A frequently pursued question in many domains within psychology is whether a set of features more than a simple sum of its parts (e.g., Asch, 1969; 1964). Within the domain of human memory, this question has taken the 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) or MINERVA (Hintzman, 1988) have adopted the <u>co-occurrence</u> assumption and represent an association as a concatenation of the two vectors representing each of the two singletons. Models including the theory of distributed associative memory (TODAM; Murdock 1982; 1997) and CHARM (Metcalfe-Eich, 1985) assume an <u>emergent</u> representation and model it as a third vector that is independent of either of the vectors representing the singletons.

Early empirical work addressing this issue focused on paired-associate learning. Many studies were developed to uncover the conditions where learning A-C interferes with previously learned pairs A-B or D-C (where 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., Postman, 1976; Greeno, James, & DaPolito, 1971; Melton & Martin, 1972). The general strategy was to understand the underlying representations by measuring interference (a strategy we adopt in the present research, albeit in different form). The hypothesis that pairs were stored as emergent configurations was supported if learning A-C did not harm memory for A-B or D-C. The competing hypothesis was that an association was simply a link or connection between two existing items in memory with the level of interference determined by the number and strength of these links. Different sets of data favored each hypothesis. For example, some found positive transfer occurs if the cues are related to one another. That is, performance for A-C is better following learning of B-C when A & B are related (e.g. Greeno, James, DaPolito, & Polson, 1978). On the other hand, negative transfer sometimes occurred when the cue was repeated with a new response unrelated to the previous response (e.g. Greeno, et al., 1971; Martin, 1968). That is, performance for A-C is worse following study of A-D 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 only marginal. Studies have found the opposite of each (i.e., McGeoch, 1942; Greeno et al, 1978) 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.

In the last several years, these issues have been studied with new paradigms, most prominently, associative recognition (AR). In AR, participants study pairs (AB, CD) and are tested with intact (AB) pairs and rearranged (AD) pairs. To be successful in this task, participants must be able to judge whether the two items occurred together. In a typical design, the familiarity of any individual item is irrelevant because both items have been studied for all test pairs. There are two primary assumptions about how a pair of items is stored in memory: the co-occurrence assumption or the emergent feature assumption, both described earlier. A few studies have tried to distinguish these approaches using the AR paradigm (e.g., Clark, Hori, & Callan, 1993; Hockley & Cristi, 1996a, 1996b; Kahana, 2002; Part I, see Clark & Gronlund, 1996 for a review of such studies).

Recently however, with the resurgence of dual process memory models, the focus has been on whether AR is carried out by a recall or familiarity based process. For example, AR requires more time for retrieval (Gronlund & Ratcliff, 1989; Nobel & Shiffrin, 2001) than SR and sometimes produces ROCs that match those generated by recall-based models (Rotello & Heit, 2000; Rotello, Macmillan, & VanTassel, 2000; Yonelinas, 1997).[1] Note that unlike free or cued recall where the response must be generated, both items of the test pair are given to the participants in AR. Thus according to many models it should be possible to simply match the test probe to memory and forego a recall/search process. In these models it is not a necessity that a search process be adopted for AR, and one might then wonder why a recall/search process is used given that it is more effortful (some possible answers were discussed by Nobel and Shiffrin, 2001 and Diller, Nobel, and Shiffrin, 2001). Whether or not a recall/search process is used for AR, it is important to consider the possibility that different representations for pairs and single items might provide a principled way to account for some findings typically attributed to the underlying retrieval processes.

Here, we continue to explore the underlying representations for associations and single items and the relationship between the two. In Part I we used the presence or absence of interference as a way to measure similarity of representation. We mixed various classes of items in one list and noted that for AR, performance is determined by the number of items within one class. 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 is determined solely by the number of studied WF pairs. Adding WW or FF pairs to the study list has no influence on WF performance (likewise for all three types of pairs). To the contrary, pair-type did not influence single item recognition (SR). These findings are not consistent with extant quantitative memory models because such models assume some overlap of representation and therefore predict between-class interference. In co-occurrence models pairs are composed of the same features as the singles from which they are composed and both must show the same pattern of interference. In models assuming emergent pair features 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.[2] Thus, we suggested modifications that could be implemented in any extant model in order to account for the pattern of data. Specifically, we adopt the following assumptions: pairs of various types are stored with dissimilar representations. However, regardless of the pair-type in which they appear, singles are stored in a fashion determined by the type of single item, rather than the pair-type. Finally, pair features are emergent in that they are not the same as the single item features.

In this article we measure similarity of representation with the presence or absence of any interaction between different stimulus types. In this approach we do not infer similarity of representation of classes of stimuli from a decrease in memory performance, but instead from any change in memory performance that depends on the stimulus type. In particular, the current set of studies adopt a 3-phase design (e.g., Criss & Shiffrin, 2004a; Maddox & Estes, 1997) in which two lists of items are studied. Some items and some pairs repeat between the study lists, but a recognition memory test is specific to the most recent list. An "old" response is to be given if a test item had been on the second list, but not if the test item had been on list one only. Items are presented on either list or both lists in such a way that we can measure the contribution of pairs of the same type compared to the contribution of pairs of a different type.

Aside from the approach in Part I, using interference to draw conclusions about the type of stored representations is not a novel strategy. For example, many have used this technique to show the existence of separate visual and verbal codes in memory. A famous example is that of Brooks (1968). He showed Ss a block letter or a sentence and asked them to report from memory whether each corner of the letter was an internal edge or whether each word of the sentence was a noun. The nature of the response was manipulated in order to measure interference. Participants either pointed to a Y or N on a sheet of paper or said "yes" or "no" aloud for each corner of the block letter or each word in the sentence. The verbal response caused most difficulty for the sentence condition and the manual response caused most difficulty for the block letter condition. Thus it seems that verbal and visual information are represented separately and the same type of information causes the greatest level of interference. We take this thesis farther and propose that different types of pairs (that overlap in verbal and visual information) are stored in a dissimilar fashion such that a memory decision is based only on the same type of pair with no contribution of different pair-types, even when the different pair-types share an identical item.

The effects in Part I were reliable but small in magnitude. Additional verification is needed because the conclusions are far reaching and the data call for revision of existing models. Here, we focus on two related goals. The first goal is to replicate these findings with a quite different and in some respects stronger experimental paradigm.

Second, we hope to better understand the relationship between the stored representation of single items and associations, when these units of analysis share the same items.

Experiments 1 & 2

Adding pairs of the same type harms AR performance, but adding pairs of a different type does not. On the contrary, SR performance is determined by the total number of single items and not affected by the relative number of each pair-type (Part I). In these studies no items repeated, so direct interference was not measured. Instead, interference, as measured by within-class list length effects, was caused by adding other pairs to the studied list. We drew strong conclusions concerning the similarity of representation and the present study attempts to verify and strengthen these conclusions through use of a quite different paradigm. In the present experiments, therefore, we use two lists at study, repeating some items and some pairs across the two lists. The participant is asked to respond "old" only to test items (pairs in Experiments 1 & 2 or singles in Experiments 3 & 4) from the second list. This design allows us to measure changes in performance that occur when items are repeated in the same type of pair (Experiment 1) compared to when items are repeated in a different type of pair (Experiment 2).

In order to confirm our previous findings, we should find an interaction such that single items repetitions affect performance only when they occur in the same pair-type but not when repetitions cross pair-types. A model for this task is needed to specify whether such an interaction should take the form of a change in discrimination, a change in bias, or both. However, one other study is informative. Dyne, Humphreys, Bain & Pike (1990) showed that repeating items between pairs (all WW pairs) within a single study list results in an increase in the probability of calling the test item "old" (P(old)) but no change in discrimination. They showed that a number of models predict this pattern because pairs with repeated items were more familiar but this was equally true for both intact and rearranged test pairs. An increase in P(old) in the present study would obviously be consistent with the Dyne et al. results.

Finding an effect of having presented a study item on an earlier list only when the repetition was in the same pair-type but not in a different pair-type would considerably strengthen the conclusions from our previous studies. Previously, we found no interference from <u>other</u> items that were presented in different pair-types. Here, we look for no contribution of an <u>identical</u> item when it was studied in a different pair-type. This study relies on the fact 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 also test single item recognition (Experiments 3 & 4) in order to compare the patterns of data between AR and SR.

In the following four studies, participants will study two lists under incidental study instructions. For Experiments 1 (AR testing) & 3 (SR testing), single items will be repeated in the same type of pair across lists. For Experiments 2 (AR testing) & 4 (SR testing), single items will be repeated in different types of pairs across lists.

Experiment 1

Methods

<u>Participants</u>

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 (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. The set of words contained 476 hard to image words of varying environmental frequency (M=18.49; range 1-245, Kucera & Francis, 1967). Any words that might describe a face, a person, or a characteristic of either were excluded.

Procedure

Participants received two study lists separated by an unfilled break of at least 120 sec. On each trial, participants had 3 sec 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 msec inter-stimulus interval (ISI). Following the final study list, participants were engaged in a 45 sec math task before beginning an unexpected memory test. The first study list contained 52 pairs of items and the second contained 60 pairs. Prior to the 72 trial test list, participants were given examples of all the 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 & 2.

Design

List 1 contained 52 WF pairs. The nature of the 60 List 2 pairs was varied between experiments. In Experiment 1, the same condition, List 2 contained all WF pairs, with an equal number of studied pairs in each of the following conditions: studied only on List 2, studied in the exact same pair in List 1 and List 2 (Lists 1 & 2 exact), studied on List 1 and 2 but in different pairs (Lists 1 & 2 re-combined). Twelve intact pairs (targets) and eight rearranged pairs (foils) were tested from each of these three conditions. 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). The other six were constructed by making a rearranged pair from items that were only presented on List 1. These foils could be called new either because they were not on List 2, because they were not presented together, or for both reasons (List 1 rearranged). Table 1 contains an example of all the test conditions.

The construction of the Lists 1 & 2 re-combined study pairs differed between-subjects. Assume the pairs AB, CD, EF, GH, etc. were studied in List 1. For one group (N=41), two resulting re-combined pairs would be AD and CB. That is, both items from two studied pairs in List 1 were re-combined to form two studied pairs in List 2. For the other group (N=40), 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. If a participant notices that an item repeats and believes this to be important, they may rehearse the previous pairing or form some sort of chain (e.g., sentence) linking the previous pair combination with the current items. If this were the case, we might expect different performance based on how the re-combined pairs were constructed. For example, in the case where AB and CD are re-combined to form AD and CB, it would be easier to form some sort of chain between the items because they are re-combined in a consistent and obvious manner. We used an incidental study design to reduce the likelihood that Ss engage in such strategies. Comparing these two groups will allow us to access the extent to which our use of incidental instructions were successful. Studied pairs always occurred side-by-side and test pairs were always presented one above the other with no relationship between the study and test position.

Results

For hit rates (HR), a 3 x 2 mixed-design analysis of variance (ANOVA) was conducted with study condition (List 2, List 1 & 2 exact, List 1 & 2 re-combined) as the within-subjects variable and group as the between-subject variable (differing only in the construction of the Lists 1 & 2 re-combined pairs). For false alarm rates (FAR), a 5 x 2 mixed-design ANOVA was conducted with study condition (List 2, List 1 & 2 exact, List 1 & 2 re-combined, List 1 intact, List 1 rearranged) as the within-subjects variable and group as the between-subject variable. There were no main effect of group (for hits $F(1, 79)=2.232$, $MSE=.100$, $p=.139$ or false alarms $F(1, 79)=0.212$, $MSE=.119$, $p=.646$) and no interactions between study condition and group (for hits $F(2, 158)=0.061$, $MSE=.020$, $p=.940$ or false alarms $F(4, 316)=0.794$, $MSE=.030$, $p=.794$), where the groups differed only in the manner in which re-combined study pairs were constructed. Thus the data are presented collapsed over this variable. It seems that our incidental study task was effective in eliminating any tendency to rehearse previous study pairs.

There was a main effect of study condition on the hit rates ($F(2, 158)=25.40$, $MSE=.001$, $p<.001$). Bonferroni adjusted post-hoc tests confirmed the order apparent in

Figure 1. Namely, the hit rate was highest for the List 1 & 2 exact condition (M=.680, SEM=.023) followed by Lists 1 & 2 re-combined 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 false alarm rates ($\underline{F}(4, 316)=16.392$, $\underline{MSE}=.030$, $\underline{p}<.001$). Post-hoc tests confirmed the apparent trends in Figure 1. False alarm rates in those two conditions where items appeared in both Lists 1 and 2 were higher than the FARs to rearranged pairs constructed from items appearing only on one list. However, there was no difference in FARs for the List 1 & 2 exact condition (M=.259, SEM=.025) and the Lists 1 & 2 re-combined condition (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 statistically the List 1 intact FAR is different from all conditions except List 1 & 2 re-combined.

Observation of the hit and false alarm rates suggests that presenting single items on both lists in different pairs of the same type does not harm discrimination relative to presenting items only on List 2. Both the hit rate and the false alarm rate for the Lists 1 & 2 re-combined condition are greater than the List 2 condition and by approximately the same amount (.075 for the HR and .093 for the FAR). Thus, in signal detection terms, it appears that there is a bias to call a test pair "old" if the items had appeared on multiple lists. However, the situation is different for the Lists 1 & 2 exact condition. Again, the hit and false alarm rates are greater than the List 2 condition, but now the magnitude of the hit rate difference (.16) is much greater than the FAR difference (.082). This seems to indicate that encoding of a pair improves when given a second opportunity, in conjunction with the previous conclusion that participants are more willing to call a pair old if its members had been studied on multiple times. Measures of discrimination confirm these observations. We computed d' from the HR and FAR of the List 1 & 2 re-combined condition, from the HR and FAR of the Lists 1 & 2 exact condition, and from the HR and FAR of the List 2 condition.[3] The resulting values are shown in Table 2. A repeated-measures ANOVA showed a main effect of condition ($F(2, 160)=5.641$, $MSE=.417$, $p=.004$). Bonferroni adjusted post-hoc tests confirmed what is described above, no difference in d' for the List 2 condition ($M=0.998$, $SEM=.096$) and the List 1 & 2 re-combined condition ($M=0.937$, $SEM=.095$) but superior discrimination for the Lists 1& 2 exact condition ($M=1.258$, $SEM=.103$).

In summary, when items being tested in AR were repeated in the same type of pair in a prior study list, participants are more willing to call the test pair old, regardless of its actual status. If the test items were presented in an identical pair on both lists, this still exists in addition to an even higher hit rate, indicating improved encoding for twice presented pairs. Intact foils from List 1 have a very high false alarm rate, indicating a lack of perfect list discrimination, as is typical in single item recognition. Experiment 2 contrasts these findings to the case where items are repeated in different types of pairs.

Experiment 2

Methods

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 Experiment 1.

Procedure

The procedure was identical to Experiment 1.

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 due to the constraint that a pair could not be repeated exactly on List 2 given that List 2 did not contain the same type of pair as those presented in List 1. An equal number of study pairs came from each of the following conditions: studied only on List 2 and studied on List 1 and 2 but in different pairs (Lists 1 & 2 re-combined). Eighteen intact pairs and twelve rearranged pairs were tested 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. These foils could be called new either because they were not on List 2 or because they were not presented together or for both reasons (List 1 rearranged). Table 3 contains an example of all the test conditions.

The construction of the List 2 re-combined pairs differed between-subjects in exactly the same fashion as Experiment 1. For one group (N=32), all four items from any two studied pairs in List 1 were re-combined to form two study pairs for List 2. For the other group (N=26), an item from one studied pair could be paired with any item from another pair except and there were no cases of type of pairing just described.

Results

For the hit rates, a 2 x 2 mixed-design ANOVA was conducted with study condition (List 2 and List 1 & 2 re-combined) as the within-subjects variable and group as the between-subject variable. For FARs, a 3 x 2 mixed-design ANOVA was conducted with study condition (List 2, List 1 & 2 re-combined, and List 1 rearranged) as the within-subjects variable and group as the between-subjects variable. There were no significant main effects of group (for hits $\underline{F}(1, 56)=0.557$, $\underline{MSE}=.079$, $\underline{p}=.459$ or false alarms $\underline{F}(1, 56)=0.070$, $\underline{MSE}=.043$, $\underline{p}=.792$) and no significant interactions between study condition and group (for hits $\underline{F}(1, 56)=1.029$, $\underline{MSE}=.015$, $\underline{p}=.315$ or false alarms $\underline{F}(2, 112)=0.363$, $\underline{MSE}=.014$, $\underline{p}=.696$), where the groups differed only in the manner in which re-combined pairs were constructed. Thus the data are presented collapsed over this variable. Again, this may indicate that the incidental nature of the study task was effective, as Ss do not seem to be rehearsing previous pairings of the study items.

Hits and false alarm rates are pictured in Figure 2. The hit rates for the List 2 condition (M=.533, SEM=.028) and Lists 1 & 2 re-combined conditions (M=.570, SEM=.029) did not differ, $\underline{F}(1, 56)=2.649$, $\underline{MSE}=.015$, $\underline{p}=.109$. False alarm rates differed by study condition ($\underline{F}(2, 112)=10.701$, $\underline{MSE}=.015$, $\underline{p}<.001$). According to Bonferroni adjusted post-hoc tests, both conditions that appeared on List 2 had similar FARs (for List 1 & 2 re-combined 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). Given that there is no difference between HR or FAR to the List 1 & 2 re-combined and the List 2 conditions, we expect no difference in discrimination. Indeed, an analysis of d', reported in Table 2, showed no significant differences between the conditions ($\underline{F}(1, 57)=1.305$, $\underline{MSE}=.238$, $\underline{p}=.258$).

Comparison of Experiments 1 & 2

We have noted that when items are repeated in the same type of pair, Ss 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 no such contribution of the repetitions (i.e., Experiment 2). In order to draw stronger conclusions about this interaction, we now directly compare the corresponding conditions of the two experiments. A 2 x 2 x 2 mixed designs ANOVA was computed with experimental group as the between-subjects factor and condition (List 1 & 2 re-combined 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 experiment and condition such that P(old) is greater for the List 1 & 2 re-combined condition relative to the List 2 condition only when the items repeat in the same type of pair and no such effect when items repetitions occur in a different pair-type. Indeed, we do find this interaction between experimental group and condition $\underline{F}(1, 137)=9.803$, $\underline{MSE}=.022$, $\underline{p}=.002$. In addition, we find main effects such that P(old) was higher to targets than foils and the to the List 1 & 2 re-combined than List 2 items ($\underline{F}(1, 137)=306.999$, $\underline{MSE}=.048$, $\underline{p}<.001$ and $\underline{F}(1, 137)=16.458$, $\underline{MSE}=.022$, $\underline{p}<.001$, respectively). No other interactions were significant, nor was there a main effect of experimental group (all $\underline{F}$'s<1 and $\underline{p}$'s<.334).

Discussion of Experiments 1 & 2

The important difference between the same condition (Experiments 1) and the different condition (Experiment 2) is the type of pair presented in 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 while List 2 contained WF pairs.

This manipulation led to a different pattern of results. For both targets and foils, we see that having presented the items on a previous list in the same type of pairs (i.e., the List 1 & 2 re-combined condition) induces participants to call the test items "old" more often compared to the case where items were presented only on List 2. When items are repeated in a different type of pair, it is as if List 1 never occurred, as we see little change in performance. This is consistent with our previous findings showing a list length effect within, but not between pair-type and with models where different pair-types are coded with dissimilar and non-overlapping representations.

These findings are also consistent with the Dyne et al (1993) study showing an increase in bias but no change in discrimination when items are repeated in multiple pairs during study. Dyne et al emphasized the importance of eliminating backward rehearsals of repeated items. Given that we found no difference in performance when re-combined pairs were constructed in different manners, we are fairly confident that the incidental task was successful in discouraging displaced rehearsals. We now turn to experiments that test single item recognition under study conditions identical to those used here.

Experiments 3 & 4

In the representations proposed in Part I, emergent pair features differ for the various types of pairs and are separate from the representation for single items but the representations for single items are similar regardless of the type of pair in which they were studied. This is based in part on a study showing that while performance in AR is determined by the number of studied items of the type being tested, performance in SR is not affected by manipulations of the relative number of different pair-types. That is, SR is determined by the total length of the list and not the number of pairs of the relevant type. In our discussions in Part I, 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 where 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 (Light & Carter-Sobell, 1970; Tulving & Thompson, 1973). The co-occurrence models necessarily predict the same qualitative pattern of results for AR and SR because the pair features and single item features are identical. Models assuming an emergent set of features for associations allows for different qualitative patterns because the tasks are based on different sets of features (associative or single item) containing different information. The results of the following studies will allow us to better understand which of these assumptions is most appropriate as we continue to develop the REM model. The study conditions of these experiments are identical to Experiments 1 & 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

Methods

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 Experiment 1.

Procedure

The procedure was identical to Experiment 1 with the exception that participants are instructed to say "old" to items studied on List 2 and reject all others.

Design

Just as in Experiment 1, List 1 contained 52 WF pairs and List 2 contained 60 WF pairs. An equal number of pairs in List 2 contained items that were studied only on List 2, studied on both lists but as members of different pairs (Lists 1 & 2 re-combined), and that were studied in identical pairs on both lists (Lists 1 & 2 exact). Because we found no difference between the two methods for re-combining study pairs described in the earlier experiments, we only used one method for Experiments 3 & 4. In particular, we chose the method of randomly selecting a re-combination. The test list consisted of 120 trials, half words and half faces. The foils consisted of 6 faces from List 1, 6 words from List 1 and 48 items (half faces and half words) that were not previously studied. The targets consisted of an equal number of words and faces from each of the three conditions described above.

Results

A 2 x 3 (item type and study condition) repeated-measures ANOVA was conducted on the hit rates. Hit rates to faces were higher than HRs to words, $\underline{F}(1, 24)=5.932$, $\underline{MSE}=.042$, $\underline{p}=.023$. There was a main effect of study condition, $\underline{F}(2, 48)=15.880$, $\underline{MSE}=.026$, $\underline{p}<.001$) and no interaction between the two $\underline{F}(2, 48)=0.092$, $\underline{MSE}=.020$, $\underline{p}=.912$. Post-hoc analysis confirm what is shown in Figure 3, Panel A;

namely hit rates for items presented in Lists 1 & 2 (M=.606, SEM=.037 for the exact condition and M=.642, SEM=.037 for the re-combined condition) are both greater than the hit rate for items presented only on List 2 (M=.470, SEM=.035) but do not differ from one another.

A 2 x 2 (item type and foil type) repeated-measures ANOVA conducted on the FARs showed higher FARs to faces compared to 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 4.

Experiment 4

Methods

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 Experiment 2.

Procedure

The procedure was identical to Experiment 2 with the exception that participants are instructed to say "old" to items studied on List 2 and reject all others.

Design

Just as in Experiment 2, List 1 contained 26 WW and 26 FF pairs and List 2 contained 60 WF pairs. Half of the List 2 pairs contained items that were studied only on List 2 and the other half contained items that were studied on both lists but in different pairs (Lists 1 & 2 re-combined). The test list consisted of 120 trials, half words and half faces. The foils consisted of 6 faces from List 1, 6 words from List 1 and 48 items (half faces and half words) that were not previously studied. The targets consisted of an equal number of words and faces from each of the two conditions presented in List 2.

Results

A 2 x 2 (item type and study condition) ANOVA was conducted on the hit rates. Hit rates to items presented on both Lists 1 & 2 ($M=.667$, $SEM=.035$) was much higher than the hit rate to items presented only in List 2 ($M=.535$, $SEM=.031$), $F(1, 24)=22.246$, $MSE=.020$, $p<.001$. There was no main effect of item type or 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 x 2 (item type and foil type) ANOVA was conducted on the FARs. FARs to items presented in 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 hits and false alarms collapsed over item type. For a breakdown by item type, see Table 5.

Comparison of Experiments 3 & 4

Observation of Figure 3 along with the individual statistical analyses from Experiments 3 & 4 both indicate that we find the same pattern of results for SR regardless of whether the single items are repeated in the same or different type of pair. In particular, we find that items studied on List 1 are more likely to be called "old" than targets studied only on List 2 or foils that have never been studied. Here we directly compare the corresponding conditions of the two experiments. A 2 x 2 x 2 mixed designs ANOVA was computed for HRs and for FARs. In both cases, experimental group was the between-subjects factor and condition (List 2 and List 1 & 2 re-combined) and item type (faces and words) were the within-subject factors. HRs do not differ for words and faces ($F(1, 48)=.825$, $MSE=.029$, $p=.368$) but those items studied on both lists have a higher hit rate than List 2 targets ($F(1, 48)=43.477$, $MSE=.027$, $p<.001$). We find an interaction between item type and experiment due to the higher HR for faces than words in Experiment 3 but not 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 was no main effect of experimental group and no other interactions approached significance (all F's<1.097 and all p's<.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 List 1 ($F(1, 48)=68.940$, $MSE=.029$, $p<.001$). There was no main effect of experimental group and no interactions of group with any of the other variables (all F's< 1.752 and all p's<.192). Thus, as expected given the individual analyses, we find the same pattern of data for single item recognition regardless of whether item repetitions occurs in the same or different type of pair.

Discussion of Experiments 3 & 4

In summary, we find 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, where only repetitions of the same type shape performance. This pattern suggests that single items are stored in a similar fashion regardless of the type of pair in which they were encoded. We now describe the model that successfully fit the patterns of data found in Experiments 1-4.

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

The REM model was first proposed for single item recognition by Shiffrin & Steyvers (1997). In Part I, we propose modifications that would allow the model to accommodate the different classes of pairs used here. The general idea is that encoding of WF, FF, and WW pairs leads to dissimilar associative information, despite the fact that they share single items of the same type. We make no specific claims about the encoding mechanism that results in these dissimilar representations, only that some combination of the incidental task, instructions, and stimulus materials causes storage of functionally separate representations. We discussed a number of ways to implement dissimilarity between pair-types in a vector model like REM. One approach utilizes a subset of vector positions in common for all pairs that identify the pair-type (the type-code assumption); another posits non-overlapping vector positions assigned to different pair-types. Choosing between these two possibilities is not necessarily interesting. The general point is simply that some combination of stimulus type and encoding strategies result in extremely dissimilar representations for the three pair-types use in this research. We adopt the type-code assumption in the REM instantiation that follows.

First consider the information stored during study of a pair. We assume that for each study trial, a memory trace is stored containing the following sets of features: single item features identifying each item, associative features relating the two items in some unique way, the current context, and a type-code identifying what type of stimulus was just encoded.[3] The information stored about single items and their resulting association are independent of one another, an assumption borrowed from Murdock (1982) and supported by various studies (Dosher & Rosedale, 1997; Hockley & Cristi, 1996b; Clark & Gronlund, 1996; Kahana, 2002; Part I). 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. If a study pair contains repeated items, the same set of features are used to store both single item repetitions (though they may differ due to the inherent error in storage as described later). Thus the representations from which features are selected to store single items are identical regardless of pair-type. However, the associative features for the two study pairs (assuming they are different as in the re-combined condition) are similar only by chance. For simplicity, the context features are assumed to be constant throughout a single study list and change with some probability between successive lists. The features identifying the type of stimulus are assumed to be identical within type but similar only by chance between types. Thus it is possible, though unlikely, for between-class traces to contribute to the memory decision.

REM assumes that feature values have differing environmental base rates, with each feature independently generated according to a geometric distribution with parameter g as follows

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

where V refers to the feature being generated and j refers to some specific feature value, j=1, 2, 3, and so forth. We assume 15 features represent each part of the vector as described above, and each value is drawn from a geometric distribution with parameter g=.40. At study only some of the features from these representations are stored in the vector, with un-stored features coded as zeros indicating a lack of information about those features. In particular, a feature is stored with some probability, u, otherwise a zero is stored. Given that a feature is stored, the correct value is copied with some probability, c=.90, otherwise a random value is drawn from the geometric distribution and stored. In our simulations, the number of features and the values of c and g are fixed as specified above. The value u was adjusted to produce a good fit to the overall level of performance. It seemed natural to assume that associative features require more effort than single item features to generate and store. We therefore fit one value of u_{assoc}=.20 for associative features and another value of u=.32 for all other features. Because both study lists are relatively similar both in time and in the encoding task being performed, 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. All of these parameters and parameter values are identical for single item and associative recognition for Experiments 1-4.

Next we turn to the retrieval assumptions. We start by noting that for the current simulations, we used the simplifying assumption that a type-code exists and that a probe with a given type-code selects from memory all traces with that type-code, and no others rather than implement that assumption. Consider single item recognition. SR proceeds just as described in Criss & Shiffrin (2004a), which differs slightly from the original REM retrieval rules. For a given test item, the system probes memory with item features and the List 2 context features (because the task requires an "old" response only to List 2 items). These are compared to all memory traces 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, λ_{iI} is calculated for a memory trace *i* in the following way:

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

where *niq* is the number of non-zero 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. The term before the product represents discounting due to mismatching features between the probe and memory trace. The term after the product represents the positive evidence gained from the matching features. Because this likelihood is based on item features alone, it is termed λ_{iI}. In parallel, memory is probed with the relevant context features. Another likelihood value, λ_{iC}, is calculated based on comparing the context features to the contents of memory using Equation 2. For this comparison, *niq* is the number of non-zero mismatching context features and *njim* is the number of matching context features with the value *j*. The term λ_{iI} gives the degree to which the memory trace matches the probe in item information and the term λ_{iC} gives the degree to which it matches in context information. A single item recognition test requires that the probe match both item and context information, so the two likelihood values must be combined appropriately. We combine the two using a weighting parameter, α , that allows the system to differentially weight item or context information as follows

$$\Phi = \frac{1}{N} \sum_{i} \left[\alpha \lambda_{iI}^{-1} + (1-\alpha) \lambda_{iC}^{-1} \right]^{-1} , \qquad (3)$$

where N is the number of memory traces contributing to the decision and Φ is the odds that the test item was studied in the relevant context. If the odds is greater than some criterion, the item is called "old" otherwise it is called "new." If the value of α equals 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, indicating an equal weighting of item and context information. For single item recognition testing, it was necessary to use a non-optimal criterion of 1.5 indicating that Ss were conservative and only claimed an item was studied if it was very familiar. This makes some sense considering the high similarity between the two study contexts causes familiarity of all studied items to increase without differentiating much between those items that were on List 2 (and should be called "old") and those items that were on List 1 (and should be called "new").

Associative recognition differs from item recognition in the type of memory probes employed. Figure 4 illustrates the decision process described here. First, as stipulated earlier, the type-code is used to probe memory, isolating comparisons to those memory traces that contain a pair of the same type as the probe; this is not a part of the calculations, but simply assumed. Thus WF probes exclusively activate WF traces from the study list(s). Next, associative features are compared those associative features stored in traces contained in the activated set. The comparison are done as for the single item features and context features via Equation 2, where *niq* is the number of non-zero mismatching associative features and *njim* is the number of matching associative features with the value *j*. The resulting likelihood ratios, λ_{iA}, give the degree to which each memory trace matches the test probe in associative features.

Others have made the assumption that an associative memory probe does not contain context features. This conclusion was based on various studies showing no forgetting for pair relative to singles over a moderate range of study-test lags (Hockley, 1992; Hockley & Consoli, 1999). We adopt the same assumption, so context is ignored and the associative activations calculated are combined by the following equation into an odds, $\phi_{\text{associative}}$:

$$\Phi_{\text{associative}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}} \lambda_{iA} \quad , \tag{4}$$

Even for associative tests, we assume that the familiarity of single items is automatically generated using Equations 2 and 3, resulting in an odds value, ϕ_{item}, for each individual item. In a typical AR experiment, this single item information is not diagnostic since all single items were presented on the study list. Thus in these typical studies, using this automatically generated single item familiarity to augment the associative decision is not helpful and only can hurt performance. Hence our model for such AR studies ignores single item information. However, there are AR studies in which consideration of single item familiarity is adaptive. In Part I, we pointed out that under certain conditions, Ss may adopt a strategy of using single item familiarity to help reject AR 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 & Wixted (2001). In their paradigm, Ss were tested with intact and rearranged test pairs as well as foils constructed from two unstudied items. They found that the FAR to unstudied foils fell below the FAR to rearranged foils (among other manipulations and findings). We suggested that their participants could have 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 adopt the same retrieval strategy here.

Assume therefore that participants in an AR test probe with the type-code and has available for decisions $\phi_{associative}$ for the test pair, and ϕ_{item} for each individual item in the test pair. Note that the familiarity of the single items is based on the match between the probe and those traces in the set activated by the type-code, and hence will not include single item traces that were stored as part of pair-types differing from that tested. If both single items are judged to be new (using for each a default criterion of 1 for the odds), the pair is called "new." If either one of the items (or both) is judged to be old, the decision is completely determined by the familiarity of the associative features. We found the best fitting criterion for the associative decision to be 0.9, indicating that Ss were somewhat generous in calling pairs "old," perhaps sensible given that any pair containing items from List 1 were presumably rejected based on the single item probes.

No parameters (excluding the various criteria just described) were allowed to vary between the same and different groups or between the groups given associative or single item testing. Fits were not completely optimized, but the fitting process was stopped when a reasonable fit was found (the predictions shown are based on 500 simulations). The fits, shown as white circles in Figures 1, 2, and 3 are quite remarkable particularly in light of the limited parameter search and the fact that all four groups were fit with the same parameter values. Observation of the graphs indicates no major deviations between the predicted and observed values. Most impressive is that the model used here is exactly the model suggested in Part I before the present data were collected. This model was suggested as a plausible way to accommodate both our Part 1 data and Kelley & Wixted's (2001) data. This model is applied here to a quite different paradigm with no additional assumptions and yet fits with high accuracy.

Good fit notwithstanding, we admit that variants of the specifics of this model are possible. We believe, however, that all 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 the pair-type is repeated. One might instantiate this idea without pair-type-codes, but instead with different pair-types represented by dissimilar and/or non-overlapping types of features. Until further studies provide additional constraints, choosing between such variants is probably a matter of taste.

We should also consider further the assumption that context plays no role in AR. One could justifiably argue that this assumption is a bit extreme. It seems likely that context features are part of the AR probe, as with any probe, but they may not play as important a role as in single item recognition for various reasons including limited capacity. This line of thinking does not imply, however, that one can simply include context features in the associative probe and eliminate the use of single item familiarity in the current model. Without single item familiarity, the model predicts approximately equivalent FARs whenever a rearranged pair is tested. 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, the model predicts an approximately equal FAR for rearranged pairs regardless of the number of times or lists on which the individual single items were studied. For the same reason, the model predicts approximately equivalent P(old) for all pairs that were presented only on List 2 (i.e., the List 2 and Lists 1 & 2 re-combined conditions) regardless of whether the single items comprising the pair were studied once or twice. These predictions do not agree with our data. The main benefit of adding context features to the associative probe is that it will serve to reduce the List 1 intact FAR that would otherwise act as a target, as the association had been stored during the experiment. Though it may seem 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.

The assumption that context is not used with an associative probe is testable in future studies: 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.

The double checking mechanism introduced here is similar, on the surface, to a model introduced by Dosher (1984; Dosher & Rosedale, 1991). Participants studied semantically related and unrelated pairs for an AR test. Semantically related pairs were easier to learn but required longer decision times. Some participants showed pronounced FARs to related pairs when the decision process was brief, but the high FARs disappeared with longer decision times. The authors suggested that that the participants invoked a secondary process late in decision making that is used to reject semantically related foils and double check before accepting semantically related targets. This example serves to illustrate another situation in which participants are induced to adopt a strategy of using additional probes before making an AR decision.

As mentioned earlier, several studies measuring ROCs and decision time suggest that AR may be carried out via a search process (e.g., Rotello, Macmillan, & VanTassel, 2000; Nobel & Shiffrin, 2001). Why then do we adopt a familiarity-based model? In part, we do so because the present results do not allow us to distinguish a familiarity based decision from one based on an elaborative search process. A search model would greatly increase model complexity, requiring many additional parameters and processes, but would not provide additional insight into the current set of data. Thus we saw little point in pursuing such a complex approach. We note also that studies used to argue for a search model tend to assume (explicitly or implicitly) that a pair is stored with the same set of features as the single items. This assumption may be partially responsible for the conclusions drawn about the retrieval processes. For example, the Nobel & Shiffrin (2001) and Gronlund & Ratcliff (1989) studies show longer decision times for AR than SR. This could be explained by a recall strategy as suggested by Diller, Nobel, & Shiffrin (2001) or as suggested by Gronlund & Ratcliff by time to generate the associative features used in the probe. The recall/familiarity distinction cannot be resolved by our present studies, so we adopted a familiarity based model for simplicity. Most critically, either type of model would require different representational similarity for different pair-types, the main point of this article.

No extant models are able to account for the results of this study and Part I without additional assumptions. The co-occurrence models cannot account for the current set of data because we find qualitatively different patterns of interference for AR and SR. 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 memory effects in AR only when the item repetitions occur in the same type of pair, composite models would also require an assumption that similarity differs between different pair-types.

Experiment 5

All studies prior to Experiments 1-4 constructed foils by combining items from the same class of items. For example a WF foil would contain a face from one WF pair and a word from another WF pair. Regardless of the source of the component parts, this foil probe would be compared to pairs of the same type (i.e., WF traces). But what would determine the similarity to such traces? Consider the model implemented above. In this model, the type-code would limit comparisons to pairs of the same type (i.e., WF pairs). Because rearranged pairs are only randomly similar to the traces from which they were generated, the model would predict an equal false alarm rate for foils constructed within class and foils constructed between class (e.g., a WF foil constructed from a face from a FF pair and a word from a WW pair). Even if single items entered the decision, though presumably they would not, the model would make the same predictions. If, however, if the single item encoding is biased by the pair-type (as suggested by the alternate model described in the previous section) and single item familiarity is automatically brought to the decision, then FARs to cross-class foils should be lower than foils constructed in the typical manner. Though this experiment was not designed as a strict test of the two model conceptions, the results could provide useful information. If we find a difference in FARs between cross-class and within-class foils then we can be somewhat confident that single item information is contributing to the decision (even though it is not diagnostic) and that the encoding of single item information is dependent on the type of pair in which it was studied.

Methods

Participants

Forty-three Indiana University undergraduates participated in the experiment in exchange for either partial course credit or $7.00 per hour.

Materials

Faces and words were randomly selected from the same set as Experiment 1.

Procedure

Participants studied 144 pairs, equally divided between WF, FF, and WW pairs. Each pair was studied for a total of 3 sec with a 500 msec ISI during which the participant was required to answer the question, "Do these two items go together?" Following study, participants engaged in a distracter math task for 45 sec. The test consisted of 32 WF pairs, half intact and half rearranged. Of the rearranged test pairs, half were constructed from items studied in WF pairs and half were constructed by taking one word from a WW pair and one face from a FF pair. As always, pairs were studied side-by-side and tested one above the other such that the test arrangement was not predictive of the study arrangement. In summary there is one independent variable manipulated within-subject, namely the type of rearranged pair. Because there is a single hit rate, false alarm rates are the only dependent variable of interest.

Results & Discussion

The hit rate was .569 (SEM=.027). The false alarm rate was slightly greater for within-class foils (M=.230, SEM=.024) than cross-class foils (M=.206, SEM=.028). However, this small numerical difference failed to reach statistical significance, $F(1, 42)=1.02$, MSE=.011, $p=.319$. We can be fairly sure that single item information, biased toward the type of pair in which it appears, is not contributing to the AR decision. This is not particularly surprising given the results of Part I and had been the working assumption until Experiments 1-4 required otherwise. This study does not allow any further conclusions regarding the proposed models.

Summary

In a paradigm where 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 changes (an overall increase in P(old) in this case) when items were repeated across lists in the same type of pair, but not 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 the AR decision. In particular, if the stored features for both single items indicate that neither were studied on the relevant study list, the pair is called new regardless of the familiarity of the associative features. If it is determined that either (or both) of the single items were in fact studied on the recent list, then the AR decision is based strictly on the familiarity of the associative features.

Footnotes

[1] Though note that some studies (e.g., Kelley & Wixted, 2001) report curvilinear ROCs for AR and SR.

[2] Note that the most recent version of TODAM (Murdock, 1997) assumes context is not used for AR decision and the memory vector is not reset at the beginning of the experiment. In combination, these assumptions produce no forgetting for pairs. Earlier versions of TODAM (Murdock, 1982) did produce forgetting due to 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 simultaneously predict both patterns as we find.

[3] When computing d', any value of 0 was replaced with 1/2N and any value of 1 was replaced with 1-1/2N where N is the number of observations in that condition.

[4] This actually transforms the vector used in REM into a matrix, but this is purely cosmetic change.

Table 1. An Example of each Study and Test Condition for Experiment 1.

Study List 1	Study List 2	Test Pair	Condition Label
	8 cow	8 cow (target)	List 2
	9 gloomy	9 cow (foil)	
3 tea	3 bear	3 bear (target)	Lists 1 & 2 Re-combined
4 bear	7 tea	7 bear (foil)	
7 smart			
11 house	11 house	11 house (target)	Lists 1 & 2 Exact
12 brain	12 brain	12 house (foil)	
5 room		5 believe (foil)	List 1 Rearranged
6 believe			
17 ape		17 ape (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 illustrated here simply to conserve space).

Table 2. Discrimination for Associative Recognition in Experiments 1 & 2.

	Experiment 1 (Same)	Experiment 2 (Different)
	d-prime	
List 2	0.998 (.096)	0.924 (.091)
Lists 1 & 2 Recombined	0.937 (.095)	1.028 (.099)
Lists 1 & 2 Exact	1.258 (.103)	

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

Table 3. An example of each Study and Test Condition for Experiment 2.

Study List 1	Study List 2	Test Pair	Condition Label
	7 tree	7 tree (target)	List 2
	8 truth	8 tree (foil)	
1 2	4 car	4 car (target)	Lists 1 & 2 Re-combined
3 4	1 house	1 car (foil)	
car house			
5 6		5 table (foil)	List 1 Rearranged
hat table			

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

Table 4. Hit 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 & 2 Re-combined	.688 (.042)	.596 (.044)
Lists 1 & 2 Exact	.648 (.040)	.564 (.049)
	False Alarm Rate	
List 1	.320 (.047)	.240 (.036)
New	.160 (.025)	.062 (.014)

Note: Standard errors around the mean are listed in parentheses.

Table 5. Hit and False alarm Rates as a Function of Item Type for Single Item Recognition in Experiment 4.

	Faces	Words
Condition	Hit Rate	
List 2	.512 (.039)	.557 (.037)
Lists 1 & 2 Re-combined	.653 (.040)	.680 (.037)
	False Alarm Rate	
List 1	.367 (.050)	.340 (.050)
New	.157 (.029)	.083 (.020)

Note: Standard errors around the mean are listed in parentheses.

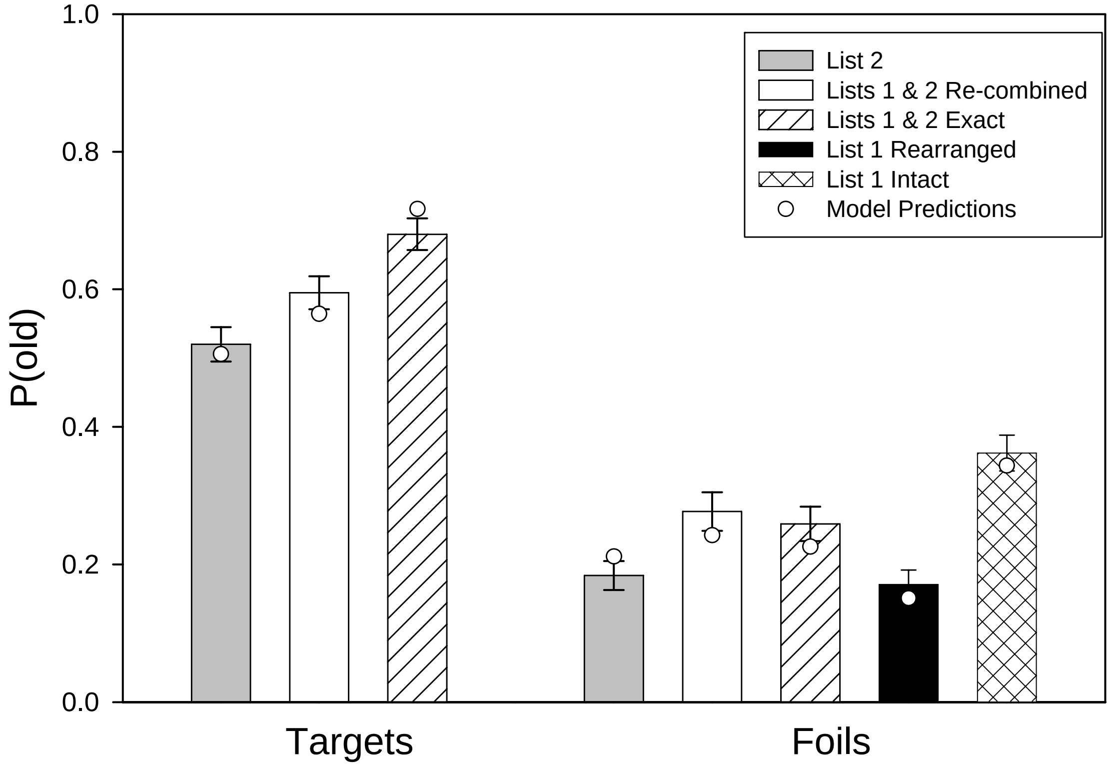
Figure Captions

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

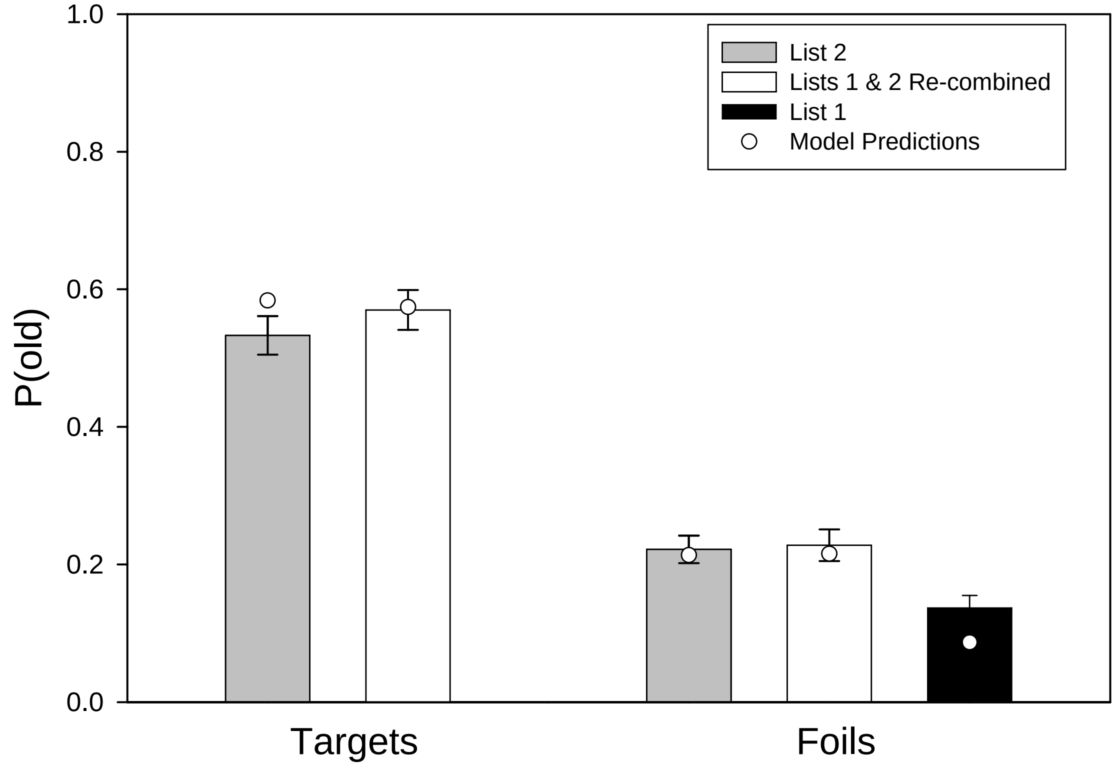
Figure 2. The probability of calling a test item old (P(old)) as a function of the type of test pair in Experiment 2. Error bars represent one standard error above and one below the mean. Open circles represent the fit of the modified REM model described in the text.

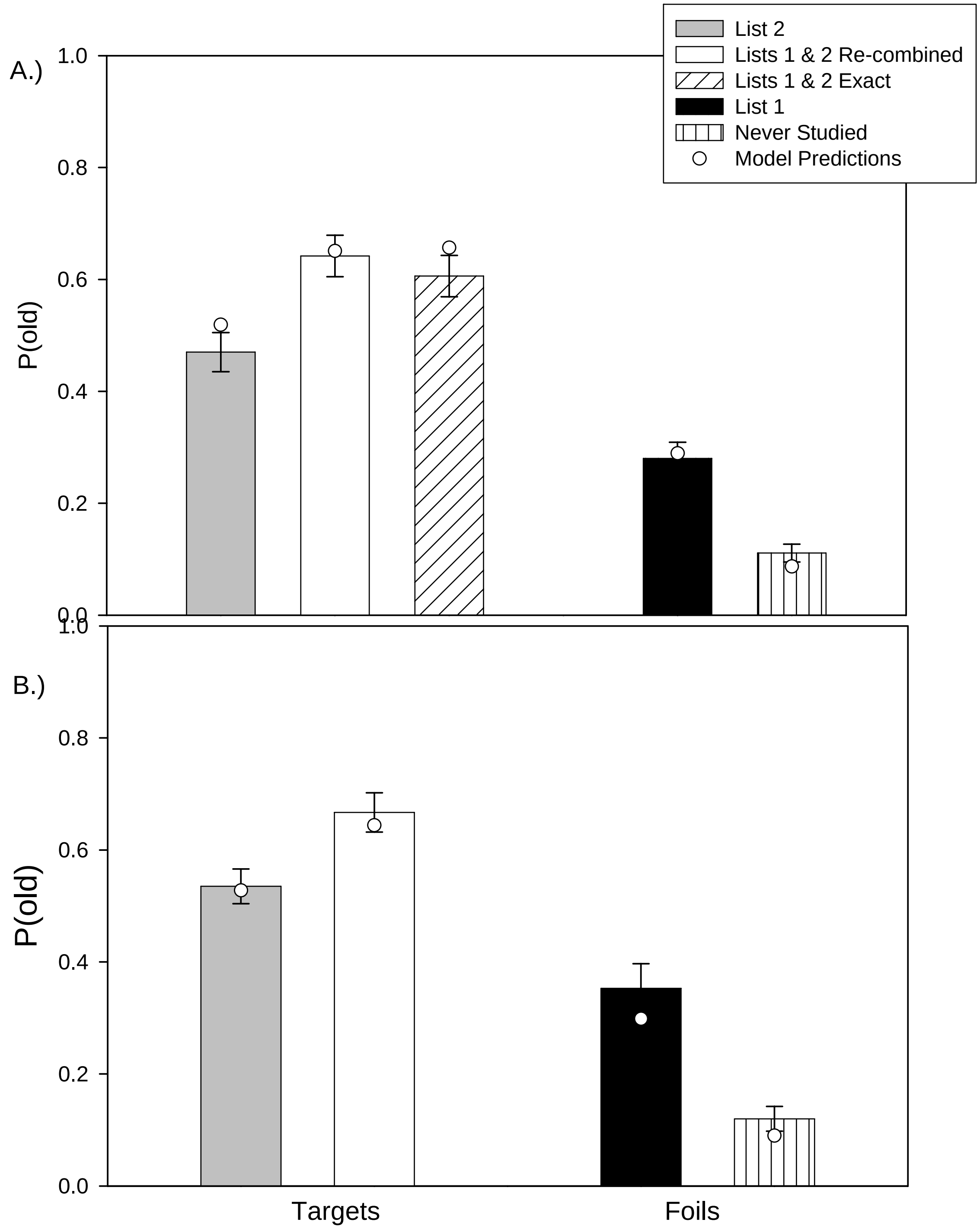
Figure 3. The probability of calling a test item old (P(old)) as a function of the type of test item. Panel A shows the data from Experiment 3 (repetitions in the same pair-type) and Panel B shows the data from Experiment 4 (repetitions in a different pair-type). Error bars represent one standard error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

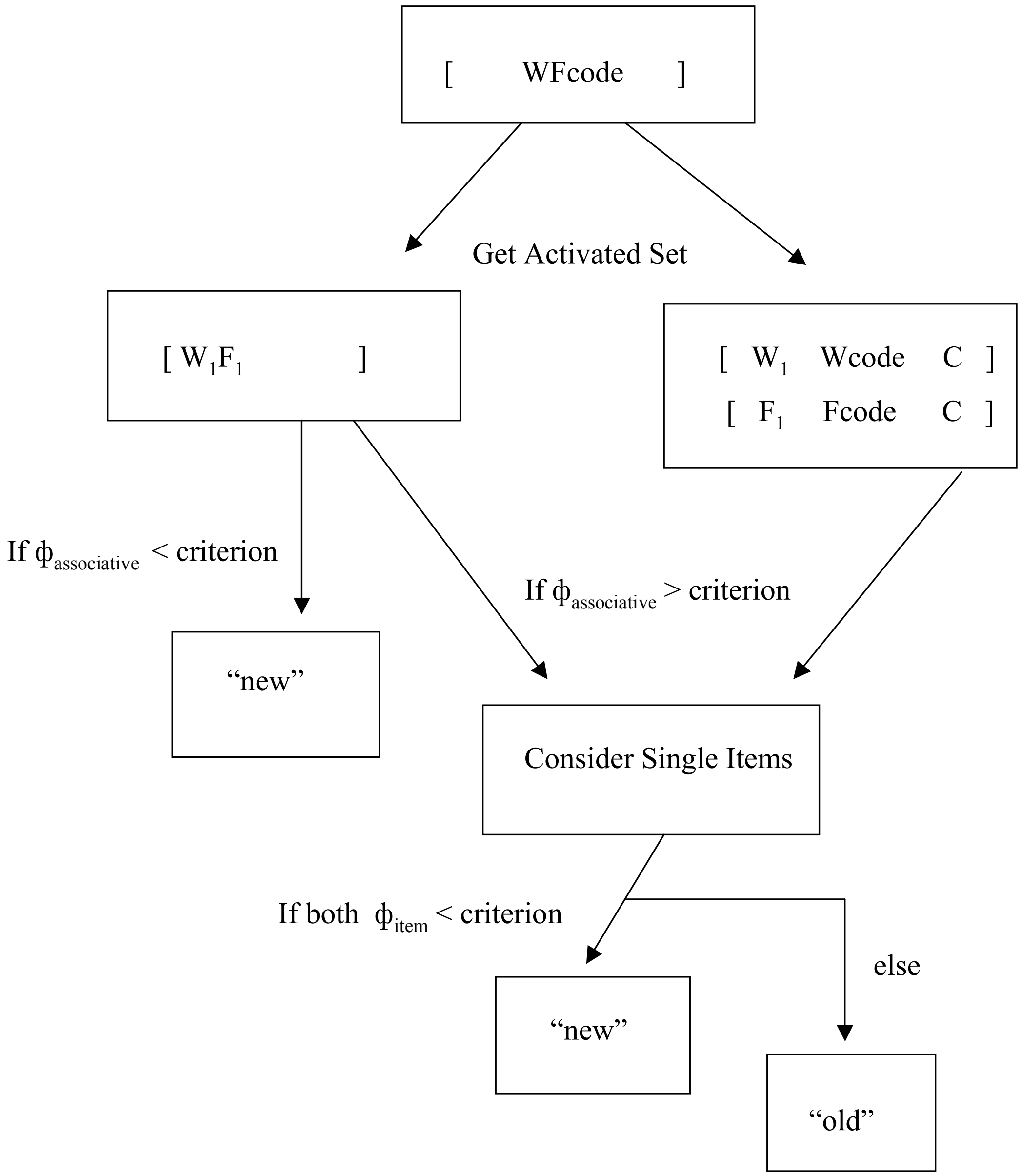
Figure 4. A schematic of the retrieval processes involved in the model used to generate the fits pictured in Figures 1, 2, and 3. See the text for a detailed description of the model.



1.0
0.8
0.6
0.4
0.2
0.0
P(old)
Targets
Foils
List 2
Lists 1 & 2 Re-combined
Lists 1 & 2 Exact
List 1 Rearranged
List 1 Intact
Model Predictions


1.0
0.8
0.6
0.4
0.2
0.0
P(old)
List 2
Lists 1 & 2 Re-combined
List 1
Model Predictions
Targets
Foils


List 2
Lists 1 & 2 Re-combined
Lists 1 & 2 Exact
List 1
Never Studied
Model Predictions
A.)
B.)
P(old)
1.0
0.8
0.6
0.4
0.2
0.0
Targets
Foils


[WFcode]
Get Activated Set
[W1F1]
[W1 Wcode C]
[F1 Fcode C]
If φassociative < criterion
If φassociative > criterion
"new"
Consider Single Items
If both φitem < criterion
else
"new"
"old"

Part III

The Role of Encoding Strategies in the Relationship Between Single Items and Associations

Parts I and II of this manuscript have demonstrated the independence of various types of pairs using experimental manipulations of list-length and list discrimination. That performance is determined by within-class but not between-class similarity is of great empirical and theoretical interest in part because this set of data cannot be accounted for by any existing model. Extant models assume similarity (in that they share comparable features) between all types of studied items whereas we find no similarity between word-face (WF), face-face (FF), and word-word (WW) pairs. A new model was designed to account for various empirical results discussed in Part I and that same model was able to beautifully predict the data from Part II. The model assumes separate features for single items and pairs that are similar to one another only by chance. For example, the features for the association generated during the study of pair AB and for the single item A are similar only by chance. Study of WF, FF, and WW pairs also results in separable representations. In Part II, they are separable by a set of features identifying the type of stimulus, much like context features identify the list on which an item was studied (though other options for separable pair features are discussed in Part I). When a single item is presented for a memory test, only single items contribute to the decision. Likewise, when tested with a pair in associative recognition (AR), only pairs of same type contribute to the decision. However, under more complicated AR paradigms where single item familiarity or context information is thought to be useful, Ss may adopt a strategy of using single item familiarity to guide their decision. Given the novelty and importance of these findings, there are many options for future research. First, we discuss a couple of possibilities followed by a set of studies that begin to address one possible future direction: the role encoding strategies play in the formation of independent representations of singles and pairs (Experiments 1 and 2) and of WF, FF, and WW pairs (Experiment 3).

One possible direction is to examine the role of verbal versus visual information. Upon hearing about this research, many suggest that the dissimilarity between WW and FF pairs is due to a difference in verbal and visual codes. FF pairs presumably result in the storage of pure visual information and WW pairs presumably result in the storage of pure verbal information. Because visual and verbal information are different in kind, they do not interfere with one another (e.g., Paivio, 1971). Following this logic, the dissimilarity of those pairs to WF pairs is due to the fact that Ss know WF pairs are a mix of verbal and visual information while WW and FF consist of only one type of information. We could use pictures of houses, for example, along with FF and WW pairs to further examine this kind of proposal. Houses are visual information, so they should perhaps interfere with FF but not WW pairs. Another approach is to require use of a verbal or visual mediator to connect the two study items. This could, perhaps, alter the type of code so that it is not strictly dependent on the stimulus. For example, study of a WW pair and a visual mediator would presumably lead to a trace containing both visual and verbal codes. Some combination of these two techniques could help us better understand the role (if any) of different codes for verbal and visual information in the finding of independence between WW, FF, and WF pairs.

Another area ripe for future research is the role of encoding strategies in the current findings. To what extent are these findings determined by either the stimulus set itself or the encoding processes acting upon the stimuli? Several studies have demonstrated the importance of encoding technique on single item recognition (SR) and associative recognition. For example, instructions to form an interactive image at study harms (Begg, 1978; McGee, 1980) or has no affect (Begg, 1978; Hockley & Cristi, 1996a) on SR relative to instructions to form separate images. However, performance in AR under the later set of instructions is considerably worse than under the former instructions. The studies just discussed address the level of performance as a function of encoding, here we begin to examine the role of encoding strategies on the type of information stored in memory. In Experiments 1 & 2, Ss study a single list containing either WF pairs alone or WF pairs and single faces. Ss are given different instructions in order to examine the encoding conditions under which the information about single items does not contribute to an AR decision. In Experiment 3, we attempt to better understand the role of encoding strategies in the dissimilarity between WF, FF, and WW pairs. We essentially replicate a subset of conditions from Experiment 3 in Part I but give Ss different study instructions.

For all analyses we use an alpha level of .05 with Bonferroni adjustments for post-hoc tests. We are primarily interested in changes in discrimination (though we also report P(old)), particularly given the risk that criteria may change between Ss groups. In order to be confident that the patterns of d_a (Macmillan & Creelman, 1991) are not due to elimination of participants with undefined values or other potential pitfalls of using the measure, we also compute and report values of d' (Green & Swets, 1966). When computing d-prime, any value of 0 or 1 was replaced with 1/2N or 1-1/2N, respectively.

Experiment 1

Recall once again that our model assumes single items do not interfere with AR decisions unless warranted by the experimental design and instructions as in Part II. That is under typical situations, such as those employed here, only associative features are used when deciding if a pair is intact or rearranged. In all prior experiments, single items were embedded in pairs. Pair-types that were not the same class as the test item did not produce interference and thus we assumed that single items did not contribute to a memory decision for a pair. For example, both WW and WF pairs contain words. If WW pairs do not affect performance for WF pairs then neither must single words. This experiment attempts to directly test this notion by including single items on the study list. Study lists will contain WF pairs, or single faces in addition to WF pairs, followed by tests of AR. According to the model, probing with a WF should activate only WF features and no features of single items, thus discrimination should increase as the number of WF pairs decreases, but the number of single faces should have no impact.

Methods

Participants

137 Indiana University undergraduates participated in the experiment in exchange for either 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 (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. The set of words contained 476 hard to image words of varying environmental frequency (M=18.49; range 1-245, Kucera & Francis, 1967). Any words that might describe a face, a person, or a characteristic of either were excluded.

Procedure

Each participant received one study-test block; the composition of the study list varied between-subjects. Each participant received a study list containing either 60 WF pairs, 40 WF pairs, or 40 WF pairs plus 20 single faces. Each trial lasted 3 sec with a 500 msec inter-stimulus interval (ISI). During study, participants were simply instructed to study the items for a later memory test with no specific task given during encoding. The length of the distracter task separating study and test varied such that the time between the first study trial and the first test trial was constant between groups. Participants in all groups received an AR test containing 40 trials, half intact and half rearranged. No test trials included faces studied as singles. In summary there was one independent variable manipulated between-subject, namely the composition of the study list.

Results & Discussion

According to the model, AR performance should be determined by the number of pairs and not affected by the presence of singles. Contrary to predictions, the level of discrimination depended on the total number of study trials rather than on the number of studied pairs, as shown in Figure 1. A one-way ANOVA established a main effect of condition ($F(2,126)=5.241$, $MSE=1.023$, $p=.007$). Post hoc analyses confirmed that the 40 WF group ($M=1.109$, $SEM=.084$) had the highest level of discrimination and there was no difference between the groups that studied 60 WF ($M=0.824$, $SEM=.057$) or 40 WF plus 20 single faces ($M=0.856$, $SEM=.060$). The 40 WF condition had a numerically higher hit rate (HR) and lower false alarm rate (FAR) than the other conditions, as can be seen in Table 1. However, the hit rate advantage was only marginally significant ($F(2,$

136)=2.509, MSE=.049, p=.085) and the false alarm rate advantage was not significant (F(2, 136)=0.822, MSE=.012, p=.442).

Why does this set of data not conform to the model predictions? It is possible that the model is simply wrong. However, the model was developed around a specific set of experiments demonstrating (albeit indirectly) that the number of single items has no bearing on performance in AR. So at least in some cases it is necessary for the model to assume that singles do not interfere with memory for associations. In Parts I and II, we always used an encoding task asking Ss to make a judgment about the pair. Given the present design, those encoding instructions were not possible. Naively, we decided to forgo giving specific instructions and simply tell Ss that their memory would be tested following the study list. Perhaps this decision was critical to the resulting pattern of data. Previous studies indicate that given such vague instructions, participants tend to prepare for a recall test (Eagle & Letier, 1964; Elias & Perfetti, 1973). Assuming the same strategy was employed here, participants likely co-rehearsed items occurring over several trials by forming sentences, stories, or by using other mnemonic devices rather than forming unique associations between items that occurred together in a single trial. This of course implies that the distinction between WF, FF, WW, and single items may be some combination of stimulus type and encoding strategies. Experiment 2 is essentially a replication of this study where Ss are given specific encoding instructions designed to encourage Ss to form associations between items within a trial but not across trials.

Experiment 2

The basic logic behind the design of this study is identical to Experiment 1 with one important exception. During study, Ss rate the ease with which they could generate a sentence containing the item(s) on the screen. Assuming Ss follow instructions, we expect to find the pattern of results predicted by the model. Namely, we expect AR discrimination to be determined by the number of pairs and not depend on the number of studied singles. The exact number of pairs and singles is different from Experiment 1 for reasons no longer relevant, but the logic is the same.

Methods

Participants

191 Indiana University undergraduates participated in the experiment in exchange for either partial course credit or $7.00 per hour.

Materials

Faces and words were randomly selected from the same set as Experiment 1.

Procedure

Each participant received a study list with either 64 WF pairs, 32 WF pairs, or 32 WF pairs plus 32 single faces. Immediately following 3 sec of study, Ss rated how easy it was to generate a sentence containing the current study item(s). Participants in all groups received 32 test trials, half intact and half rearranged. All other details are identical to Experiment 1.

Results & Discussion

Instructions to form a sentence including the item(s) on a trial lead to data more consistent with the model predictions than instructions to simply study for a later memory test. The level of discrimination depended on the number of studied pairs and was not affected by the additional studied single items, as shown in Figure 2. Unfortunately, this effect is weak and not statistically significant as measured by a one-way ANOVA ($\underline{F}$(2, 164)=.949, MSE=.427, p=.389). Both hits and false alarms were numerically greater in the condition with the fewest trials (e.g., 32 pairs) suggesting these Ss adopted a less strict criterion. This effect was only significant for hit rates ($F(2, 188)$=.4.198, MSE=.138, p=.016) but not for false alarms ($F(2, 188)$=.298, MSE=.009, p=.742). Hits, false alarms, and d-prime are reported in Table 2.

Though the expected pattern of data is present, the effect is weak and not statistically reliable. Why do we not find stronger evidence supporting the model assumption that memory traces from single items are not involved in making a memory decision about a pair? There are at least two possibilities. First, it is likely that participants vary in the degree to which they follow instructions during study. For example, the group studying WF pairs plus single faces may have begun to form redundant sentences for the single faces. It seems likely that generating a different sentence for a series of faces is a tedious task, particularly given our set of faces tended to be somewhat uninspiring. Though they do vary in race, age, and gender, the people pictured tend to be wearing a collared shirt and a smile, and are void of any bizarre hair styles, facial markings, or clothing (given that the majority are taken from yearbooks). Perhaps the variability in adopted study procedures also encouraged variability in test strategies. This is in part supported by the observation that the size of the standard error of the mean for the condition where both singles and pairs were studied is greater than the other conditions (see Table 2).

A second and perhaps more compelling reason that the manipulation did not produce a more reliable set of data is the nature of the encoding question. Ss were asked to generate a sentence involving the item(s) on each trial, for either WF pairs or a single face. For WF pairs, Ss likely generated a sentence linking the word to the face in some meaningful way. For single faces, Ss must also generate a set of words that are related to a face. These two strategies sound strikingly similar, the only difference being whether one of the words from the sentence was provided by the experimenter. Unfortunately these may have resulted (at least in some cases) in the storage of somewhat similar representations for the different types of stimuli.

Regardless of the exact reasons for the somewhat weak pattern of data, Experiments 1 and 2 are informative. What is clear from this data is that the measurable dependence or independence of different types of pairs and single items is not purely a function of stimulus type but depends on encoding. When encoding instructions emphasize a unique relationship between items presented in a single trial (e.g., in all previous studies using the "How well do these two items go together?" task), the different types of pairs and singles seem to be encoded in such a way that they do not share features and thus do not contribute to the decision for one another When no encoding instructions are provided (e.g., Experiment 1), Ss default encoding strategy is one in which singles and pairs are encoded in way that allows them to interfere during a memory test. In Experiment 2, we find a pattern somewhat in between, perhaps due to a less than optimal choice of encoding task.

Experiment 3

In previous studies showing within-class but not between-class effects, participants were given a task during encoding. In fact, in most research on AR, participants are given a task to encourage relational encoding. The majority of such instructions include either a suggestion to form a sentence or an image relating the two items. As part of a pilot study prior to those studies reported in Part I, we manipulated the encoding task and found the best overall performance when Ss were given a task where they decide how well the two items go together, thus we adopted this task for the majority of our studies. As already described, several studies have shown that the level of performance in an AR task depends on the encoding instructions. For example, studies (Hockley & Cristi, 1996a; McGee, 1980; Begg, 1978) have shown that instructions emphasizing the relationship between two study items improves AR performance while instructions emphasizing the individual items harms performance. In terms of the current findings, the question becomes whether the dissimilarity of the various pair-types depends on the encoding instructions. Experiments 1 & 2 demonstrated that the similarity between singles and pairs is at least partially dependent on encoding task. Thus, it is reasonable that we will find a parallel results for the different pair-types.

The current study is similar to Experiment 3 from Part I. Participants are given two relevant study-test blocks each containing two types of pairs. The number of studied pairs is held constant for one pair-type and is varied for the other pair-type. This design allows us to independently measure the effects of adding pairs of the same type as well as adding pairs of a different type to the study list. Instead of using the encoding task used in Part I (e.g., "How well do these two items go together?"), we simply tell Ss to study the items for a later memory test. If the ‘go together’ instructions simply encourage associative encoding in general but not a different type of encoding for the three pair-types, we expect to see the same pattern here (perhaps with slightly worse performance overall). Namely, we would expect within but not between class length effects. However, if the independence of the pair-types depends on both the stimulus class and the encoding task then performance may be based on the total list length, not the number of pairs of the same type.

Methods

Participants

96 Indiana University undergraduates participated in the experiment in exchange for either partial course credit or $7.00 per hour.

Materials

Faces and words were randomly selected from the same set as Experiment 1.

Procedure

The design is a 2 x 2 x 2 mixed design with condition (constant or varied) and list (short or long list) varied within-subject and pair-type assignment varied between-subject. Participants are given a total of 3 study-test blocks with 45 sec separating study and test. Each study pair was presented for 3 sec separated by a 500 msec ISI. The first block is practice to familiarize participants with the task and the resulting data is not analyzed or presented here. For the practice list, Ss studied a total of 20 pairs (10 of each type). The practice test list contained a total of 20 pairs, 5 intact and 5 rearranged pairs from each pair-type. The remaining two study-test blocks contain two conditions of interest. In the constant condition, 20 pairs of that type are presented on each study list. In the varied condition, one study list contains 20 pairs of that type and the other study list contains 40 pairs of that type. In particular, Group A studied the following two lists in addition to the practice list: List A (the short list) contained 20 WF and 20 FF pairs and List B (the long list) contained 20 WF and 40 FF pairs. For this group, WF pairs are assigned to the constant condition and FF are assigned to the varied condition. Group B received lists with FF as the constant pair-type and WF as the varied pair-type. The order of the short and long study lists was chosen randomly for each Ss. To control study-test lag, each study list was constructed such that only the first 40 pairs contributed to the test list and no others. After each study list, Ss receive 40 test pairs, 10 intact and 10 rearranged pairs from each of the pair-types and are instructed to accept intact pairs and reject rearranged pairs. Note that the two lists of interest are constructed in a similar fashion as the corresponding conditions in Experiment 3, Part I. One non-critical difference is the absolute number of study pairs (here, it is 20 vs. 40 and in the Part I study it was 10 vs. 20). Of particular importance, in the present design we do not ask Ss to engage in the go together study task; they are simply instructed to study for a memory test.

Results & Discussion

In a previous study, we showed that for a given type of test pair, performance is determined by the number of studied pairs the same type (Experiment 3 of Part I). Adding pairs of a different type to the study list did not change discrimination. If the dissimilarity of the various pair-types is due solely to stimulus differences then we should see the same pattern of data here.

In fact, we did not replicate those findings. Performance was determined by the total list length rather than the length of the pair-type. We performed a 2 x 2 x 2 mixed design ANOVA with condition (constant or varied) and list (short or long) as within-subjects factors and assignment of pair-type to condition as the between-group factor. First consider the different groups of participants (i.e., whether FF served as varied or constant condition). For all dependent measures there will be an interaction between group and condition because for one group WF pairs will be the constant condition and the other group will have WF pairs in the varied condition. This interaction, therefore only tells us that performance varies between WF and FF pairs and this is not particularly enlightening. There were no other interactions or main effects involving group, so the remaining discussion ignores this variable (Table 3 contains a full list of the dependent measures for both groups).

To replicate the previous finding showing that performance depends on the number of within class pairs but not on the number of between class pairs, we should find an interaction between list and condition such that performance for the varied condition is greater in the short compared to long list but performance for the constant condition does not differ between lists. This pattern of data is exactly what we found previously and is shown in Figure 3, Panel A. In the current experiment, we find no evidence for this interaction by observation of the data or by statistical tests. Instead, we find that performance is better for the short list for than the long list for both the constant and the varied conditions.

First consider d_a, pictured in Figure 3, Panel B, we find better performance for short lists compared to long lists and better performance for the constant condition than the varied condition ($F(1, 35)=3.107$, $MSE=.431$, $p=.087$; $F(1, 35)=9.003$, $MSE=.259$, $p=.005$, respectively) but no interaction between the two, $F(1, 35)=1.688$, $MSE=.343$, $p=.202$. There is no reason that discrimination in the constant condition should exceed that in the varied condition. In fact, it appears that the opposite was true for the prior experiment. Thus, we do not consider this difference meaningful or discuss it further. Despite the statistical analyses indicating otherwise, observation of Figure 3, Panel B may suggest that we do find the expected pattern of data, however in this case the reported values of d_a are particularly noisy due to elimination of about 60% of the participants due to undefined values. The pattern of data and statistical tests for HR, FAR (shown in Figure 4) and d-prime (shown in Table 3) confirm that in fact, performance was based on the total list length and not the number of within class pairs.

The list length effect is mainly attributable to differences in the FARs. We find greater false alarm rates for the long list than the short list ($F(1, 94)=5.556$, $MSE=.022$, $p=.020$), no main effect of condition ($F(1, 94)=1.245$, $MSE=.028$, $p=.267$) and no interaction ($F(1, 94)=0.168$, $MSE=.023$, $p=.683$). There was no difference between HRs for the long or short list ($F(1, 94)=0.381$, $MSE=.023$, $p=.539$) or the constant versus varied conditions ($F(1, 94)=1.385$, $MSE=.040$, $p=.242$) and no interaction between the two variables ($F(1, 94)=0.236$, $MSE=.025$, $p=.629$).

The primary difference between this study and our previous study was the set of encoding instructions. Much like Experiments 1 & 2, when Ss are not given specific instructions they tend to encode items (different pair-types here, singles and pairs in Experiments 1 & 2) in a similar fashion. We hypothesize two ways in which this is accomplished. First, they tend to co-rehearse items from nearby trials which muddles the difference between types of stimuli. Second, they may tend to use some generic encoding strategy such as rote repetition which does not allow for dissimilarity between types of stimuli to emerge. We can be certain that the dissimilarity observed in previous studies is not purely a function of the stimulus itself and it is dependent on encoding strategies. This raises the interesting possibility that one could potentially find independence between different WW pairs, given certain encoding constraints. For example, suppose Ss studied WW pairs with very different encoding tasks. It may be possible to demonstrate independence within the class of WW pairs given such a design.

What are the implications from this study for future research? Depending on the goals, one should carefully decide whether to use generic instructions or instructions designed to emphasize the relationship between the two items (e.g., how well do these items go together). The exact task employed during study is not so important as recognition and discussion that any pattern of results may be dependent on the study task and not a general property of memory, as authors would often prefer to claim.

General Discussion

Since at least Atkinson & Shiffrin's (1968) designation of control strategies as "those processes that are not permanent features of memory, but are instead transient phenomena under the control of the subject; their appearance depends on such factors as instructional set, the experimental task, and the past history of the subject" the field has acknowledged that encoding strategies play a major role in memory. Much of the early research involved showing different levels of performance depending on the type of encoding (e.g., Craik & Lockhart, 1972). For example, several studies demonstrated that under a variety of conditions 'deep' or semantic tasks (i.e., generate a sentence for the study word) resulted in better performance than 'shallow' or surface-level tasks (i.e., state how many times the letter E appears in the study word). Today, manipulations such as these are often used to improve performance. However, little research has addressed the mechanisms of why or how encoding tasks lead to different levels of performance.

Of greater interest here is not whether various encoding methods lead to different levels of performance, but whether tasks lead to storage of qualitatively different types of information. For example, we have recently demonstrated that one of the most robust findings in the recognition memory literature, the word-frequency effect (WFE), is dependent on the encoding task. The WFE is defined as a higher HR and lower FAR for words of low normative frequency compared to words of high normative frequency. In Criss & Shiffrin (2004b), we demonstrated that the HR portion of this effect is entirely dependent on encoding task. In particular, we used a relatively large number and variety of encoding tasks (mixed within a list or between lists) and found the typical finding for only two of the tasks: generic instructions to study for a later test and a task asking whether or not the word contains any distinctive letters. For all other tasks, we found no difference in HR as a function of word frequency. Note that this finding held true over the whole range of levels of performance. We explained our findings in terms of the type of features stored as a function of encoding strategies. Relevant to the current research is the idea that different types of information may be encoded depending on the task demands.

The Criss & Shiffrin (2004b) studies are a particularly compelling example of what can happen when encoding issues are addressed as we showed that the WFE, deemed a "regularity of recognition memory," is heavily dependent on encoding. In the current manuscript, we demonstrated that the type of information stored during study of a pair is determined by the encoding strategy. Under some situations, single items and pairs are stored in a way that allows them to interfere with one another and under other study instructions, they are stored with independent representations.

Another line of research is more relevant to the present studies. Several studies have shown that emphasizing information about individual items harms AR performance relative to a study instructions emphasizing the relation between items (Begg, 1978; McGee, 1980; Hockley & Cristi, 1996a). Such studies are generally used to support the assumption that associative information goes beyond just the information about the individual items. The present studies fit nicely in this literature as we show different qualitative patterns of results depending on encoding instructions. Our conclusions, however, are more specific than previous studies because we also show independence (or dependence depending on encoding instructions) of items and associative features as well as between different types of pairs.

In these studies, we simply demonstrate that encoding instructions can result in varying levels of dependence between singles and pairs and between different classes of pairs. Future work should address the mechanisms behind such an effect. Ultimately, we would like a process model of typical encoding strategies as well as the conditions under which each is employed.

Table 1. Hit Rates, False Alarm Rates, and d-prime for the Groups in Experiment 1.

	40 Pairs & 20 Faces	40 Pairs	60 Pairs
Hit Rate	.551 (.019)	.616 (.022)	.578 (.021)
False Alarm Rate	.230 (.018)	.203 (.016)	.233 (.019)
d-prime	0.941 (.065)	1.229 (.095)	1.008 (.076)

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

Table 2. Hit Rates, False Alarm Rates, and d-prime for the Groups in Experiment 2.

	32 Pairs & 32 Faces	32 Pairs	64 Pairs
Hit Rate	.622 (.026)	.680 (.022)	.588 (.020)
False Alarm Rate	.238 (.025)	.260 (.023)	.239 (.019)
d-prime	1.188 (.100)	1.264 (.114)	1.046 (.101)

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

Table 3. Hit Rates, False Alarm Rates, d_a, and d-prime as a Function of Study Condition for each Group in Experiment 3.

	HR	FAR	d_a	d-prime
		Group A		
Short List				
Varied (FF)	.583 (.026)	.272 (.028)	0.491 (.155)	0.914 (.114)
Constant (WF)	.741 (.030)	.211 (.028)	0.982 (.160)	1.626 (.133)
Long List				
Varied (FF)	.550 (.025)	.324 (.028)	0.320 (.139)	0.645 (.119)
Constant (WF)	.748 (.027)	.257 (.028)	0.977 (.156)	1.493 (.125)
		Group B		
Short List				
Varied (WF)	.702 (.025)	.246 (.027)	0.997 (.143)	1.372 (.109)
Constant (FF)	.576 (.029)	.256 (.026)	0.759 (.148)	0.959 (.127)
Long List				
Varied (WF)	.700 (.024)	.252 (.027)	0.535 (.128)	1.330 (.114)
Constant (FF)	.566 (.026)	.294 (.027)	0.633 (.144)	0.796 (.120)

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

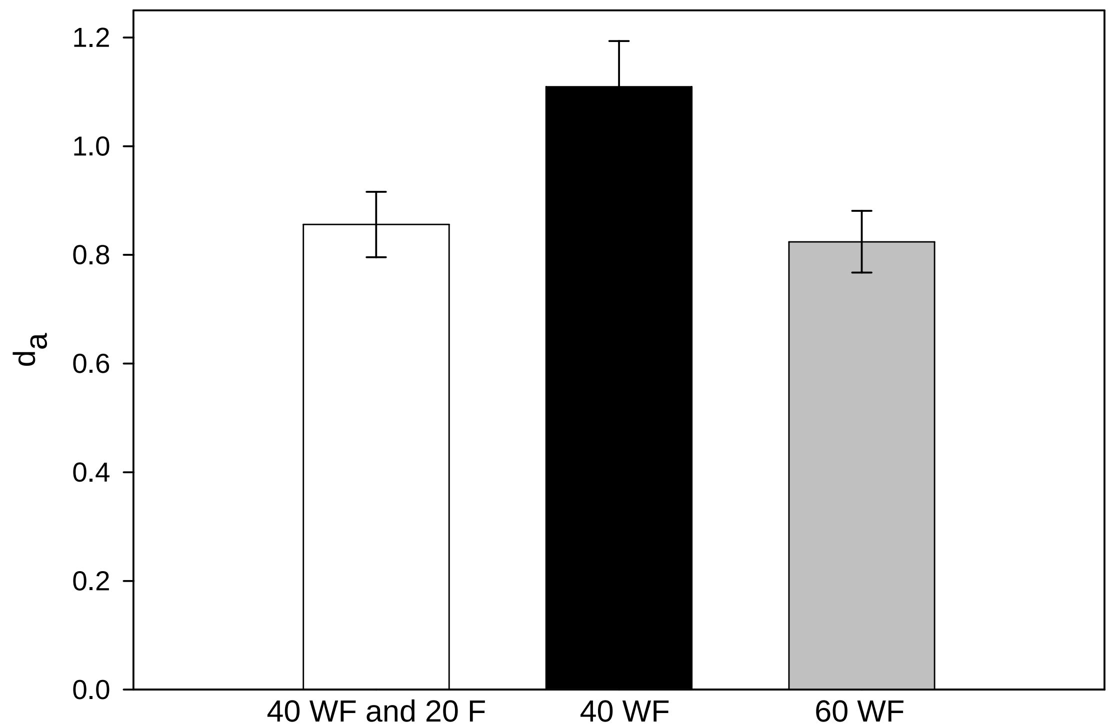
Figure Captions

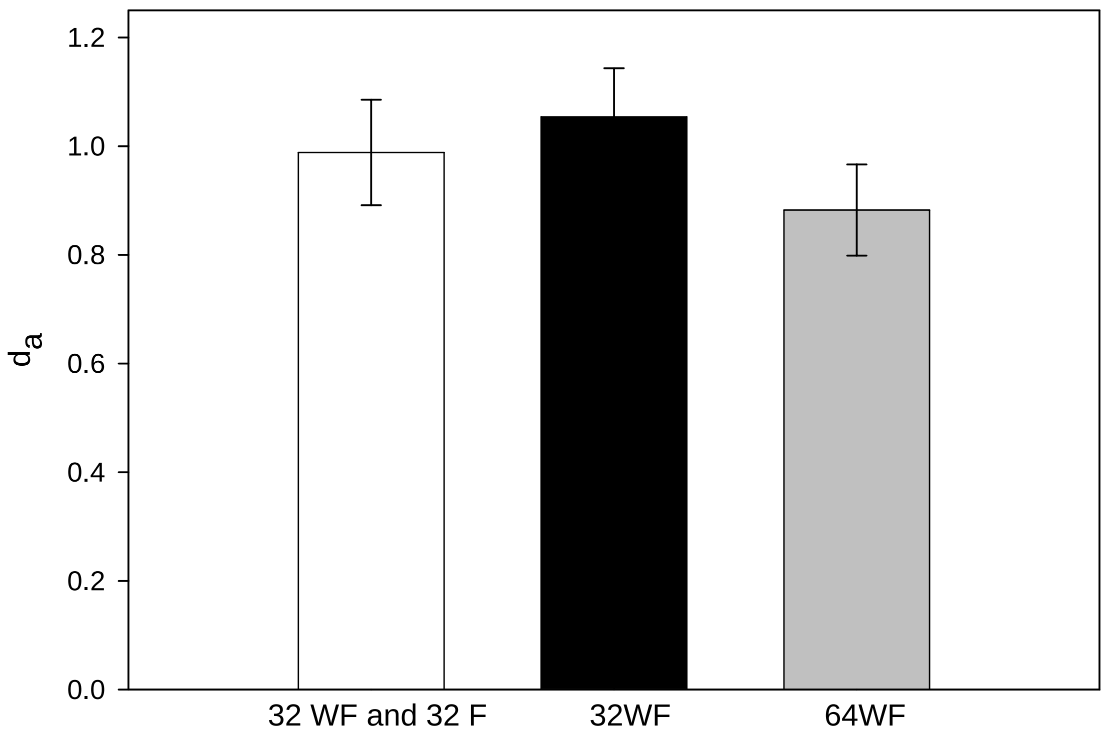
Figure 1. Discrimination as a function of study condition in Experiment 1. Error bars represent one standard error above and one below the mean.

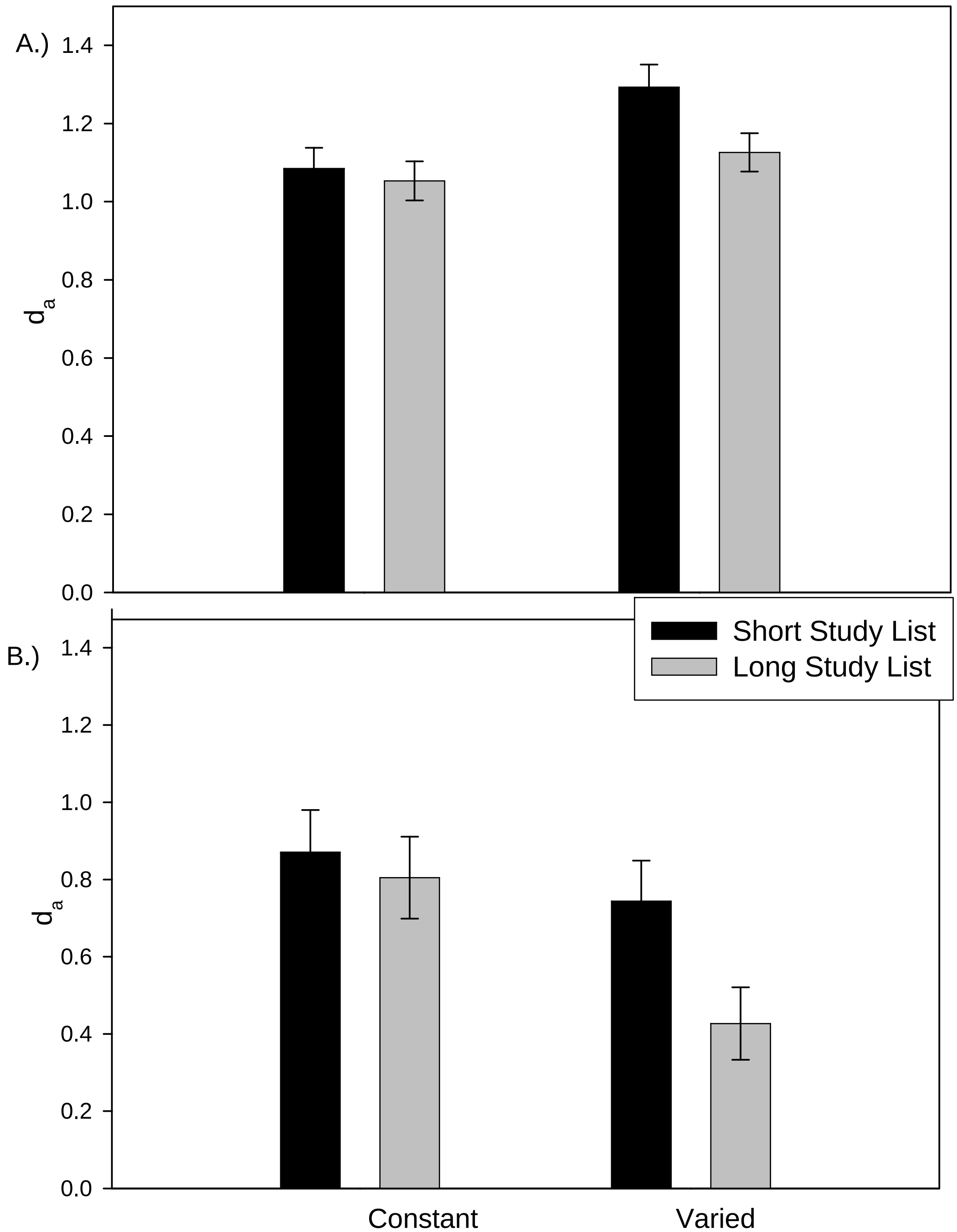
Figure 2. Discrimination as a function of study condition in Experiment 2. Error bars represent one standard error above and one below the mean.

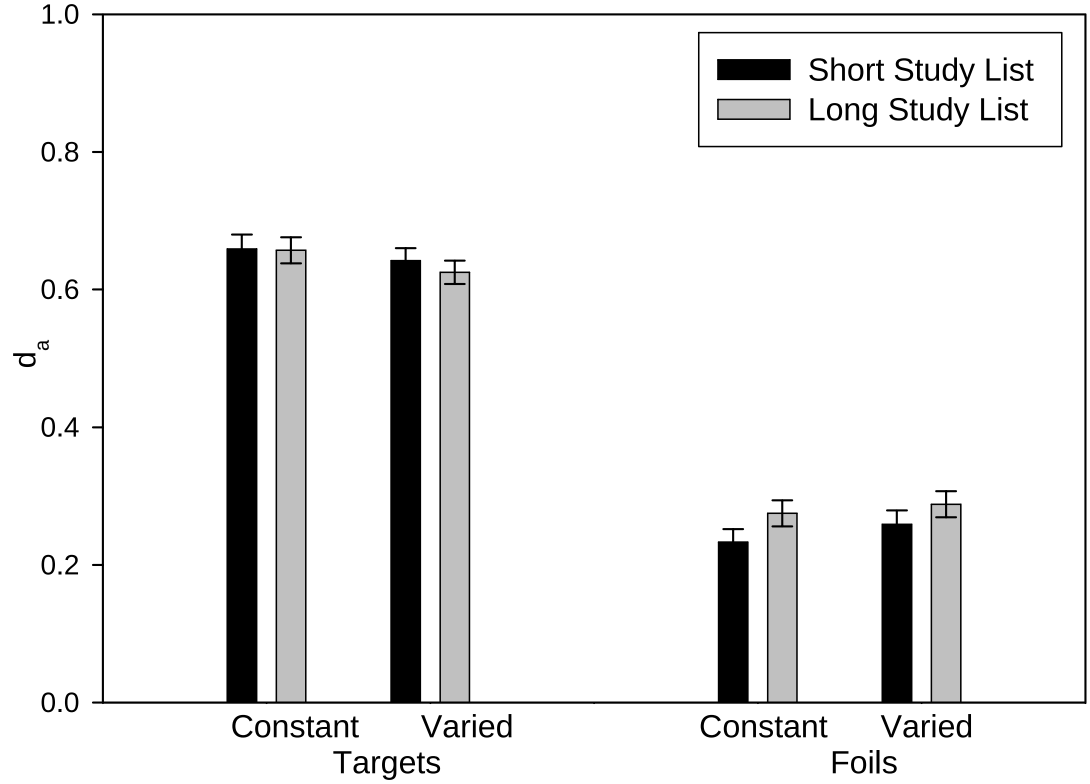
Figure 3. Panel A is a subset of the conditions in Experiment 3 of Part I. Discrimination is plotted as a function of list and condition. Panel B shows the comparable data set from Experiment 3. In both cases, error bars represent one standard error above and one below the mean.

Figure 4. The probability of calling a test item old (P(old)) as a function of list and condition for Experiment 3, collapsed over Ss group. Error bars represent one standard error above and one below the mean.


1.2
1.0
0.8
0.6
0.4
0.2
0.0
d_a
40 WF and 20 F
40 WF
60 WF


1.2
1.0
0.8
0.6
0.4
0.2
0.0
d_a
32 WF and 32 F
32WF
64WF


A.)
B.)
d_a
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Short Study List
Long Study List
Constant
Varied


1.0
0.8
0.6
0.4
0.2
0.0
d_a
Short Study List
Long Study List
Constant
Varied
Targets
Constant
Varied
Foils

References

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. Journal of Experimental Psychology: General, 128(2), 186-197.

Anderson, M. C., & Neely, J. H. (Eds.). (1996). Interference and inhibition in memory retrieval. San Diego, CA: Academic Press, Inc.

Asch, S.E. (1964). The Process of Free Recall. In: Cognition: Theory, Research, Promise (Eds. Scheerer, C.) New York: Harper & Row., pp. 79-88.

Asch, S.E. (1969). A reformulation of the problem of associations. American Psychologist, 24(2), 92-102.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), The Psychology of Learning and Motivation: Advances in Research and Theory (Vol. 2, pp. 89-195). New York: Academic Press.

A. T. & T, Cambridge (1994). The Olivetti Research Database of Faces [electronic database]. Available FTP: Hostname: uk.research.att.com Directory: pub/data/att_faces.tar.Z.

Begg, I. (1978). Imagery and organization in memory: Instructional effects. Memory & Cognition, 6(2), 174-183.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. Canadian Journal of Psychology, 22, 349-369.

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. Memory & Cognition, 20(5), 580-598.

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(4), 871-881.

Clark, S. E., & Hori, A. (1995). List length and overlap effects in forced-choice associative recognition. Memory & Cognition, 23(4), 456-461.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. Psychonomic Bulletin & Review, 3(1), 37-60.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior, 11, 671-684.

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

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.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. Psychological Review, 108(2), 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(2), 414-435.

Dosher, B.A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. Cognitive Psychology, 16(4), 519-555.

Dosher, B.A., & Rosedale, G.S. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. Journal of Memory & Language, 30(2), 125-160.

Dosher, B. A., & Rosedale, G. S. (1997). Configural processing in memory retrieval: Multiple cues and ensemble representations. Cognitive Psychology, 33(3), 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(5), 813-824.

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. Journal of Experimental Psychology, 68, 58-63.

Elias, C., & Perfetti, C. (1973). Encoding task and recognition memory: The importance of semantic encoding. Journal of Experimental Psychology, 99, 151-156

Gillund, G., & Shiffrin, R. M. (1981). Free recall of complex pictures and abstract words. Journal of Verbal Learning & Verbal Behavior, 20(5), 575-592.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. Psychological Review, 19(1), 1-67.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. Oxford, England: John Wiley.

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(4), 331-345.

Greeno, J.G., James, C.T., DaPolito, F. & Polson, F.G. (1978). Associative Learning: A Cognitive Analysis. New Jersey: Prentice-Hall, Inc.

Griffiths, T.L., & Steyvers, M. (2003). Prediction and semantic association. In: Advances in Neural Information Processing Systems, 15.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. Memory & Cognition, 15(5), 846-858.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review, 94(5), 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, & Cognition, 18(6), 1321-1330.

Hockley, W. E., & Cristi, C. (1996a). Tests of encoding tradeoffs between item and associative information. Memory & Cognition, 24(2), 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(6), 796-811.

Hockley, W. E. & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. Memory & Cognition, 27(4), 657-664.

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.

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.

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(3), 701-722.

Kucera, & Francis, W. N. (1967). Computational Analysis of Present-Day American English. Providence: 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(4), 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(1), 83-97.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York, NY, US: Cambridge University Press.

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

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. Journal of Experimental Psychology: Learning, Memory, & Cognition, 28(2), 380-387.

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. Psychological Review, 75(5), 421-441.

McClelland, J. L., & Chappell, M. (1999). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. Psychological Review, 105(4), 734-760.

McCloskey, M., & Bigler, K. (1980). Focused memory search in fact retrieval. Memory & Cognition, 8(3), 253-264.

McGee, R. (1980). Imagery and recognition memory: The effects of relational organization. Memory & Cognition, 8(5), 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.

Metcalfe-Eich, J. M. (1985). Levels of Processing, Encoding Specificity, Elaboration, and CHARM. Psychological Review, 92(1), 1-38.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. Psychological Review, 89(6), 609-626.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). Psychological Review, 104(4), 839-862.

Murdock, B.B. & Kahana, M. J. (1993). Analysis of the list-strength effect. Journal of Experimental Psychology: Learning, Memory, & Cognition, 19(3), 689-697.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin, 95(1), 109-133.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. Journal of Experimental Psychology: Learning, Memory, & Cognition, 27(2), 384-413.

Paivio, A. (1971). Imagery and verbal processes. Oxford, England: Holt, Rinehart & Winston.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. Psychonomic Science, 1(5), 125-126.

Postman, L. (1976). Interference theory revisited. In: J. Brown (Ed.), Recall and recognition. Oxford, England: John Wiley & Sons.

Raaijmakers, 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(2), 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(4), 447-472.

Rock, I. and Ceraso, J. (1964). Toward a Cognitive Theory of Associative Learning. IN: Cognition: Theory, Research, Promise (Eds. Scheerer, C.) New York: Harper & Row., pp. 110-146.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. Memory & Cognition, 28(6), 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(4), 763-797.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM--retrieving effectively from memory. Psychonomic Bulletin & Review, 4(2), 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.

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(2), 267-287.

Sohn, M.-H., & Anderson, J. R. (2001). Task preparation and task repetition: Two-component model of task switching. Journal of Experimental Psychology: General, 130(4), 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(4), 442-449.

Steyvers, M. (2001). Modeling semantic and orthographic similarity effects on memory for individual words. Dissertation Abstracts International: Section B: The Sciences & Engineering (Indiana University), 61(11-B).

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. Journal of Verbal Learning & Verbal Behavior, 5(4), 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(6), 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), 1-43.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. Memory & Cognition, 25(6), 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(4), 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(5606), 577-580.

Curriculum Vitae

Education

- Indiana University - joint Ph.D. in Cognitive Psychology and Cognitive Science with a certificate in Mathematical Modeling, (advisor: Richard M. Shiffrin)
- Miami University - B.A., magna cum laude, 1997, Psychology/Neuroscience

Awards and Honors

- College of Arts and Sciences Dissertation Year Research Fellowship, 2003-2004
- National Science Foundation Graduate Student Fellowship, 1999-2002
- Fellow, McDonnell Foundation Summer Institute in Cognitive Neuroscience, Dartmouth College, 2001
- I.U. Cognitive Science Supplemental Scholarship. 1998; 1999; 2000.

Publications

- Criss, A.H. and Shiffrin, R.M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis & Humphreys (2001). Psychological Review, 111(3), 800-807.
- Criss, A.H. and Shiffrin, R.M. (2004). Interactions between study task, study time, and the low frequency hit rate advantage in recognition memory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 30(4), 778-786.
- Criss, A.H. and Shiffrin, R.M. (in press). The lack of interference for pairs and single items in episodic memory. Memory & Cognition

Invited Talks

- George Washington University, 2004Conference Presentations Meeting of the Psychonomic Society: 2001, 1997

- Society for Mathematical Psychology: 2003, 2001
- Annual Interdisciplinary Conference: 2003, 2002, 2000
- Annual Summer Interdisciplinary Conference: 2004, 2003, 2002
- Society for Neuroscience: 1997
- Hoosier Mental Life: 2002, 2001, 2000

Professional Service

- Conference Co-organizer, Hoosier Mental Life 2001, Bloomington, IN.
- Graduate & Professional Student Organization, Cognitive Science Representative
- American Psychological Association Graduate Student Liaison
- Ad Hoc Reviewer for the Journal of Experimental Psychology: Learning, Memory, & Cognition, Memory & Cognition, and Conference Proceedings of the Cognitive Science Society

Courses Taught

- PSY P335 Cognitive Psychology
- PSY P211 Methods of Experimental Psychology

Courses Assisted

- PSY P326 The Psychology of Motivation
- PSY P335 Cognitive Psychology
- PSY P101 Introduction to Psychology I