ACADEMIC
PRESS

The effect of distinctive visual information on false recognition

Jason Arndt* and Lynne M. Reder

Department of Psychology, 342C Baker Hall, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

Using the false memory paradigm (Deese, 1959), recently revived by Roediger and McDermott (1995), we examined the effect on true and false recognition of presenting study items in unusual looking fonts. In one condition, each font was associated with a single study item. In a second condition, each font was presented 12 times per study list, randomly distributed across several themes. In a third condition, each font was presented 12 times in the study list, and was associated with a particular study theme. False recognition levels were lowest when there was a unique association between each font and a single study item, whereas false recognition levels were highest when all items from a theme were presented in the same font. Further, the effects of font condition on false recognition of lures maintained when font condition was manipulated within participants and lists. These results, taken together, are inconsistent with theories proposing that false recognition reduction is the product of global shifts in response strategies across conditions (e.g., Schacter, Israel, & Racine, 1999). However, perspectives highlighting the effects of memory based processes on true and false recognition provide an adequate account.
© 2002 Elsevier Science (USA). All rights reserved.

False memory has received a tremendous amount of attention in the past several years, both in popular circles as well as in the memory community. Within the memory community, the focus on false memory in list learning paradigms can be traced to a study by Roediger and McDermott (1995). Roediger and McDermott, reviving a paradigm originated by Deese (1959), presented participants with lists of study items that were all related to a single, unpresented item (referred to as the *lure item*). On a later free recall test, a memory task in which participants do not often intrude unstudied items, participants produced the lure item with approximately the same probability as study items presented in the middle serial positions of the study list (Roediger & McDermott, 1995). Further, participants were given a recognition memory test and asked to provide remember-know responses when an item was judged old. Results of the recognition memory test produced two notable findings. First, the false alarm rate of the lure items was approximately equivalent to the hit rate for studied items. Second, participants often assigned a *remember* response to lure items, indicating that they consciously recollected the presentation of the lure items on the study list, when, in fact, such items were never presented at study. These results have been subsequently replicated and extended by numerous researchers (e.g., Payne, Elie, Blackwell, & Neuschatz, 1996). The fact that participants falsely recognize and recall unstudied items with such a high probability, coupled with the fact that participants claim conscious recollection of a lure item's presentation,

---

* Corresponding author. Present address: Department of Psychology, Middlebury College, Middlebury, VT 05753. Fax: +1-412-268-2798.

E-mail address: jarndt@middlebury.edu (J. Arndt).

suggests that false recognition[1] is caused by fundamental memory processes. Thus, understanding the nature of false recognition may provide critical insights into the processes of memory.

Concurrently, theorists have become increasingly interested in the role that recognition memory decision processes play in the production of true and false recognition (Hirshman, 1995; Hirshman & Arndt, 1997; Miller & Wolford, 1999; Roediger & McDermott, 1999; Wickens & Hirshman, 2000; Wixted & Stretch, 2000). In general, decision based approaches to false recognition (e.g., Dodson & Schacter, 2001, 2002; Hirshman, 1995; Miller & Wolford, 1999; Schacter et al., 1999) claim that there are elements of memory representations that cause participants to change the basis by which they evaluate whether a test item should be considered studied. For example, increasing the precision with which study items are encoded may lead to an increase in the amount of information participants require to be retrieved from memory before endorsing a test item as old (Hirshman, 1995).

The hypothesis that false recognition differences are caused by decision processes is often contrasted with the hypothesis that differences in false recognition are due to memory based processes independent of decision processes, such as encoding or representational factors (Arndt & Hirshman, 1998; Hirshman & Arndt, 1997; Roediger & McDermott, 1999; Wickens & Hirshman, 2000; Wixted & Stretch, 2000). Memory based approaches to false recognition (e.g., Wickens & Hirshman, 2000; Wixted & Stretch, 2000) claim that the properties of the memory representations on which false recognition is based are the causal element producing changes in false recognition across experimental conditions. For example, increasing the precision with which a study item is encoded may decrease the match between an unstudied test item and items encoded in memory, thereby reducing the evidence that the test item was studied (Shiffrin & Steyvers, 1997).

A recent theoretical instantiation of how decision processes may reduce false recognition in the paradigm of Deese (1959) and Roediger and McDermott (1995) was proposed by Schacter et al. (1999) (see also Dodson & Schacter, 2001, 2002). Schacter et al. proposed that when participants are provided with salient visual information at study, they will come to expect that they should be able to retrieve detailed visual information at test. Such a retrieval strategy will cause participants to search more

rigorously at test for encoded visual information, a notion Schacter et al. referred to as a *distinctiveness heuristic*. When participants are unable to retrieve visual information, they will tend to reject the test item, producing a reduction in false recognition because unstudied items are less likely to have visual information associated with the representations supporting their recognition. Further, Schacter et al. (1999) proposed that a distinctiveness heuristic operated on a global basis, such that all test items were subjected to analysis for associated visual information in memory, and not just items for which salient visual information was re-presented at test.

Consistent with the operation of a distinctiveness heuristic, Israel and Schacter (1997) and Schacter et al. (1999) observed a reduction in false alarms when participants were shown pictures of study items (in addition to words) at encoding relative to participants that were only provided with words at study. Further, this reduction in false alarms occurred regardless of whether both pictures and words or only words were presented at test, suggesting that a general shift in decision strategy occurred as a result of the *study* of pictorial information. Schacter et al. (1999) further tested the distinctiveness heuristic hypothesis by varying encoding condition within participants and study lists. The distinctiveness heuristic hypothesis predicts that when encoding condition (picture + word vs. word) is manipulated within participants and study lists, similar levels of false recognition should occur across encoding conditions. The reason for this is simple: if a difference in decision strategy is the solitary reason that differences in false recognition are observed in an experiment, there should not be differences in the memory based factors that contribute to false recognition across conditions. Thus, if encoding condition is manipulated within participants and lists, and participants do not change their decision strategy across item types within a test list (see Stretch & Wixted, 1998 for evidence that participants maintain a single decision strategy within a test list), then the differences in false recognition reduction observed due to encoding condition should no longer be observed. Thus, by manipulating encoding condition within study lists, any amount of false recognition reduction resulting from a decision strategy shift would be equated across encoding conditions, and different levels of false recognition could only result from the properties of the memory representations on which false recognition is based. Consistent with the proposition that a distinctiveness heuristic was driving the reduction in false recognition, the manipulation of encoding condition within participants and study lists eliminated differential suppression of false recognition across word and pictorial encoding (Schacter et al., 1999; see Dodson & Schacter, 2001 for similar results with saying vs. hearing words at study).

The ubiquity of the distinctiveness heuristic as a mechanism of false recognition reduction is less clear.

---

[1] While the term false recognition technically applies to the erroneous recognition of *any* new test item, we are primarily concerned with the false recognition of unstudied items that are semantically related to a number of study items. Thus, we use this term to refer to this restricted case of false recognition phenomena.

For example, the distinctiveness heuristic does not appear to provide a comprehensive account of the reduction in false recognition observed when items are studied in the visual modality relative to the auditory modality. One might expect the distinctiveness heuristic to apply to this situation because visual information is encoded in the visual study condition, while little, if any, visual information is encoded in the auditory study condition. For this situation, the distinctiveness heuristic predicts that (1) false recognition will decrease in the visual study condition relative to the auditory study condition when study modality is manipulated between participants; (2) this reduction would occur regardless of the modality in which test items were presented; and (3) false recognition should *not* differ as a function of study modality in within-participants designs. Consistent with the first prediction, Smith and Hunt (1998) reported that participants that were presented with study items visually were considerably less likely to false alarm to lure items relative to a group of participants that were presented with study items aurally. With regard to the second prediction, Gallo, McDermott, Percer, and Roediger (2001) demonstrated that the effect of study modality on false recognition occurred on visual recognition memory tests, but not auditory recognition memory tests, a finding that is inconsistent with the expectations of the distinctiveness heuristic hypothesis. With regard to the third prediction, Gallo et al. (2001) also manipulated study modality within participants and lists, and found a similar pattern of results as they did with a between-participants manipulation (i.e., lower levels of false recognition for visual study items relative to auditory study items on visual, but not auditory, recognition memory tests), which is again inconsistent with the expectations of the distinctiveness heuristic hypothesis.

The results of Smith and Hunt (1998) and Gallo et al. (2001) are within the explanatory realm of memory-based explanations of false recognition, however. For example, one theory that seems to provide a reasonable account of the visual recognition test results of Smith and Hunt (1998) and Gallo et al. (2001) is the item specific and relational processing account (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993). According to the item specific and relational processing account, relational information in memory arises from the processing of the commonalities among study items, while item-specific information in memory arises from the processing of differences between study items. The encoding of relational information, while generally beneficial to memory performance, would tend not only to improve the recognition of study items, but also to increase the false recognition of semantically related lure items. In contrast, the encoding of item-specific information should only increase accurate recognition of study items. Thus, study conditions that are conducive to the encoding of relational information

would be expected to increase both levels of accurate recognition of study items and false recognition of lure items, while study conditions that are conducive to the encoding of item-specific information should only increase accurate recognition of study items. On the assumption that visual presentation of study items draws participants' attention to the item-specific attributes of items more than auditory presentation, this perspective can accommodate the effects of study modality on true and false recognition in visual recognition tests.

However, the item specific and relational processing theory does not provide a straightforward account of the effects of the presentation of pictorial information on true and false recognition. Specifically, this theory, as well as other memory-based explanations, would expect that the manipulation of encoding condition (picture + word vs. word) in the Schacter et al. (1999) would be unaffected by whether the manipulation was within or between participants. Further, it is not obvious how this theory could accommodate Gallo et al.'s (2001) finding that auditory recognition tests fail to show a modality effect without additional assumptions about retrieval cues interacting with item specific and relational representations in memory. Thus, while both theories succeed in explaining some elements of the effects of visual study processing on false recognition, neither provides a comprehensive account in and of itself.

The present experiments

In the present set of experiments, we further evaluate these examplars of memory-based vs. decision-based explanations of false recognition reduction. In particular, we presented study items in unusual looking fonts, allowing for the development of associations between distinctive visual information and each study item. Further, we manipulated the utility of the fonts as a retrieval cue by presenting fonts either once or 12 times, a manipulation recently demonstrated to influence recognition memory hits and false alarms (Diana, Peterson, & Reder, 2002; Reder, Donavos, & Erickson, 2002b). We included two conditions in which fonts were presented repeatedly during study. In one condition, the font was correlated with the semantics of the themes, such that all items from a particular theme were presented in the same font. In a second condition, participants were presented with items in a font that was repeated 12 times across the study list, but fonts were randomly assigned to several themes within the study list. Thus, in the latter case there was no correlation between the visual properties of the fonts and the semantics of the themes within the study list.

The distinctiveness heuristic predicts that false recognition should vary across encoding conditions according to the salience of the visual information

presented at study. Thus, in conditions where the visual information is highly salient (e.g., when fonts are presented with a single item), the distinctiveness heuristic would predict that false recognition would be least. In conditions where visual information is less salient (e.g., when a single font is utilized to present an entire theme), the distinctiveness heuristic would expect false recognition to be greatest. Critically, the distinctiveness heuristic would predict that differences in false recognition would not occur when font condition is manipulated within participants and study lists. Thus, when an independent variable is manipulated between participants, this hypothesis would propose that participants will be more likely to utilize a distinctiveness heuristic when visual information is highly salient than when it is not, allowing for differences in false recognition to be observed across the font conditions. However, when an independent variable is manipulated within participants (and lists), a distinctiveness heuristic should not be able to differentially suppress false recognition across conditions. Specifically, when font condition is manipulated within participants (and lists), the search for distinctive visual information in an effort to verify an item's presentation at study should fail for all non-studied items (e.g., lure items) regardless of the experimental condition to which their corresponding study items were assigned, producing equivalent levels of false recognition across conditions.

The item specific and relational processing account predicts a similar pattern of false recognition across conditions when font condition is manipulated between participants, albeit for different reasons. This account proposes that font condition should affect the properties of the encoded representations in memory, and that those differences will produce different levels of false recognition across conditions. Item-specific encoding would be expected to be greatest when fonts are presented with a single study item, a condition that would presumably draw participants' attention to the individual attributes of study items. Relational encoding would be expected to be greatest when fonts and themes are correlated with one another, a condition that would presumably enhance attention to, and encoding of, the commonalities among study items. Importantly, and in contrast to the prediction of the distinctiveness heuristic, this explanation would expect the patterns of false recognition to replicate across between and within participants (or between and within lists) designs, because the properties of encoding (and the resultant memory representations) would not be expected to differ across between- and within-participants designs.

To evaluate the predictions of these two explanations of the effects of the study of salient visual information on false recognition, we conducted three experiments. In our first experiment, we manipulated font condition between participants. In our second experiment, we manipulated font condition within participants and study lists. The comparison of the effects of font condition on false recognition across these two experiments should allow for discrimination between the distinctiveness heuristic and the item specific and relational processing accounts. In our third experiment, we compared performance in two of the font conditions utilized in the first two experiments with a condition where study items were presented in a common font. This third experiment was conducted to compare our results with those of the extant false recognition literature.

Finally, we also collected judgments of recognition memory phenomenology (i.e., remember-know judgments; Tulving, 1985) to assess the extent to which participants were judging items as studied based upon a sense of recollection or simply based upon a sense of familiarity. Previous studies in this paradigm (e.g., Roediger & McDermott, 1995) have found that not only do participants mistakenly judge lure items old with a high probability, but also typically claim to "remember" the lure items' presentation in the study list. Similarly, Schacter et al.'s (1999) results demonstrated that the reduction in false recognition when pictures were presented at study was associated with a reduction in the likelihood that participants judge lure items as "remembered" rather than "known" (or familiar). Thus, we evaluated the influence that font condition had on both overall levels of false recognition as well as how strongly those levels of false recognition were associated with recollection or familiarity phenomenology.

Experiments 1 and 2

Experiments 1 and 2 are essentially the same experiment, with one important difference. In Experiment 1, font condition was manipulated between participants, while in Experiment 2, font condition was manipulated within participants. Recall that the distinctiveness heuristic predicts that the effects of font condition on false recognition observed in Experiment 1 should not occur in Experiment 2. The item specific and relational processing account predicts that the effects of font condition on false recognition should replicate across Experiments 1 and 2. Next, we explicate the materials, design, and procedure for Experiment 1. At the end of the procedure section, we note the procedural differences between Experiments 1 and 2.

Method

Participants

Ninety students from Carnegie Mellon University participated to satisfy a research appreciation requirement for Experiment 1. Thirty-six students from the same population participated in Experiment 2.

Materials and design

Stimulus items were selected from the Nelson, McEvoy, and Schreiber (1998) database. Forty-eight sets of 12 items each (referred to as *themes* below) were chosen to serve as study items. Each set of 12 items was generated such that all items within a theme produced the same item in free association with a nonzero probability (range = .022–.913; range of mean associative strength for themes = .327–.664; mean associative strength of all items utilized in these experiments = .417). For example, the stimulus items *rose*, *stem*, *blossom*, *lily*, *vase*, *dandelion*, *orchid*, *tulip*, *petals*, *daisy*, *garden*, and *bloom* were chosen as study items, and all produce the lure item *flower* in free association.[2]

The present experiment utilizes a variety of unusual looking fonts to present study and test items. One hundred forty-four fonts were selected from various internet sources. Selection criteria were similar to those used for studies investigating font matching effects in recognition memory (Reder, Donavos, & Erickson, 2002b). Thus, the fonts utilized in this experiment were ones that participants were unlikely to have encountered previously, did not resemble fonts one typically sees in everyday life (e.g., Helvetica, Courier), were easily readable, and were similar in size to one another. Examples of the fonts selected for the current experiment are presented in Fig. 1.

For each participant, 24 of the 48 themes were randomly selected to serve as study materials, with the remaining 24 themes serving as distractor materials for the recognition memory tests. Within the set of 24 themes designated as study items, six themes were randomly assigned to each of four study lists, yielding study lists 72

Fig. 1. Examples of fonts.

items in length (six themes with 12 items per theme). Study items were presented blocked by theme, such that all 12 items from one theme were presented, followed by all 12 items from a second theme, and so on until the study items from the six themes comprising a given study list had been presented.

The test list for each set of study items was composed of all of the items presented at study, an equal number of items from unstudied themes (referred to as *new items* below), the lure items for each of the studied themes, and the lure items corresponding to each of the unstudied themes (referred to as *new lures* below). Thus, test lists were 156 items in length (72 study items, 72 new items, six lure items and six new lures). Order of presentation of themes within a study list, order of presentation of study items within each theme, and order of presentation of all test items, regardless of the theme to which they belonged, were randomly determined for each participant.

One independent variable was manipulated: the manner in which fonts were assigned to study items. Participants were randomly assigned to one of three between-participants conditions: the correlated font condition, the uncorrelated font condition, and the unique font condition. In the correlated font condition, six fonts were used per study list, with all of the study items in a theme presented in the same font, but fonts differed across themes. In the uncorrelated condition, six fonts were used per study list, with the assignment of study items to fonts determined randomly with the constraint that each font was presented 12 times in the study list. Thus, in the uncorrelated condition, a font was presented the same number of times in a study list as in the correlated condition, but none of the six themes in a study list was strongly associated with a particular font. Across all four study lists, 24 fonts were utilized per participant in the correlated and uncorrelated conditions.

In the unique font condition, each of the 72 study items was displayed in a different font. The number of study items across all four study lists in this experiment was greater than the number of fonts selected for this experiment. Thus, it was necessary to recycle each font once across the four study lists in the unique font condition. One set of 72 fonts was randomly selected to

---

[2] It is important to note that this form of list construction was somewhat different than is typically used in this paradigm. Specifically, these stimulus materials were constructed such that all of the study items produce the lure item in free association. However, the manner in which most other studies in this paradigm construct stimulus materials is to use the norms of Russell and Jenkins (1954) or norms based upon Russell and Jenkins (1954), (e.g., Stadler, Roediger, & McDermott, 1999; see McEvoy, Nelson, & Komatsu, 1999), for an exception. In the Russell and Jenkins norms, items used as study items in this paradigm were produced *by* the lure in free association, which does not necessarily mean that they produce the lure in free association. Despite this difference in the construction of the study materials, we did not anticipate fundamental differences in the phenomenon under study, other than to possibly increase the overall probability of false recognition because all of the study items elicit the lure item in free association, which is not necessarily a characteristic of lists based upon the Russell and Jenkins (1954) norms. To the extent that we are able to compare the levels of performance in these experiments to the prior literature (e.g., false alarms to lure items, false "remember" responses to lure items), these materials seem to produce similar results to those previously found in studies of false recognition.

display study items in the first and third study lists, and the remaining 72 fonts were utilized to display study items in the second and fourth study lists. Finally, across participants, all of the 144 fonts selected for presentation of study items were used an equal number of times (five times each) in the correlated and uncorrelated conditions, mitigating any concerns one might have over the random selection of fonts for display in the correlated and uncorrelated conditions.

At test, study items were presented in the font in which they were studied, while lure items were presented in a font used to present a thematically related study item. Thus, in the correlated font condition, each lure item was presented in the font used to present all of its corresponding theme items. In the uncorrelated font condition, lure items were presented in one of the six fonts used to present the theme items related to that lure. In the unique font condition, lure items were presented in one of the 12 fonts used to present the theme items related to that lure. Finally, fonts were assigned to new items and new lures in an analogous manner to that utilized to assign fonts to study items in each condition. Thus, all new items and new lures were presented in a font encountered at study, and the manner in which those fonts were assigned to new items and new lures depended on the font condition to which participants were assigned.

Procedure

Each participant completed four study-test cycles in which all essential aspects of the procedure were the same within each level of the between-participants variable. At study, participants were presented with a list of words in a number of different fonts and were asked to rate how appropriate each font was for the meaning of the word on a scale ranging from one (not very appropriate) to four (very appropriate). This task was intended to direct participants' attention to both the visual properties of the font as well as the semantic properties of the word. Participants were further provided with examples of a situation in which one might judge the correspondence between the visual properties of a font and the meaning of a word as appropriate, and one in which the correspondence might be judged as inappropriate. For example, participants were instructed that if the study item was "windy" and the font appeared to be italicized or was bent over towards one side as wind might do to an object, that correspondence might be judged appropriate. On the other hand, if the font appeared to be rigid or did not contain wavy lines characteristic of the depiction of "windiness" they might judge that correspondence to be inappropriate. Participants were further instructed that they were to try to remember the items, because an unspecified memory test would occur later in the experiment. They then proceeded through the study list at their own pace.

Immediately following the completion of the study task, participants were given a recognition memory task in which they were asked to discriminate between studied and unstudied items. Participants were also asked to provide judgments of their phenomenological experience for old responses using the remember—know paradigm (Tulving, 1985). Participants were instructed that a remember response was appropriate when they were able to consciously recollect the experience of an item's presentation at study, while know responses were appropriate when they were certain that a test item had been presented, but were unable to consciously recollect the experience of an item's presentation. Thus, there were three potential responses a participant could provide for each test item: *remember*, *know* or *new*. Participants were given extensive instruction on the difference between a remember experience and a know experience, and asked to provide an example of each to ensure they understood the distinction. Following completion of the first recognition memory test, participants were given a second study list, followed by a second recognition memory test. This procedure continued until participants had completed all four study-test cycles.

Experiment 2 alterations

The procedure for Experiment 2 was the same as in Experiment 1 with four exceptions. First, study lists contained 12 themes rather than six themes. Thus, the study list length was 144 items (12 items presented for each of 12 themes) and the test list length was 312 items (144 study items, 144 new items, 12 lure items, 12 new lures). Second, the independent variable (assignment of fonts to themes) was manipulated within participants and within study lists in this experiment. Thus, four of the 12 themes within each study list were presented in each of the font conditions (unique, correlated and uncorrelated). Third, participants completed two study-test trials rather than four study-test trials as had been the case in Experiment 1. Fourth, as a result of the first three alterations to the procedure, only 112 of the 144 fonts were utilized.[3] Finally, assignment of (1) fonts to experimental condition (correlated, uncorrelated, unique); (2) themes to experimental condition; (3) fonts to themes (and items in the uncorrelated and unique conditions); (4) themes to be studied or unstudied; (5) themes to serial position within the study list; (6) items to serial

---

[3] Each participant viewed 96 fonts used to display items in the unique condition (8 total themes across the two study lists × 12 fonts per theme), 8 fonts used to display items in the uncorrelated condition (4 fonts per study list presented 12 times across the different themes assigned to the uncorrelated condition; 8 fonts total across the two study lists), and 8 fonts used to display items in the correlated condition (one font presented with all of the study items in each of eight themes assigned to the correlated condition).

position within a studied theme; and (7) items to serial position within the test list were all determined randomly for each participant.

Results and discussion

For each experiment, we present our results in terms of analyses on the proportion of remember responses in each of the font conditions and recognition memory sensitivity. In general, across these three experiments, the largest effects of font condition were found on the proportion of false alarms (and false remember responses) to lure items. These differences are demonstrated below in terms of analyses of remember responses to lure items, as well as analyses of recognition memory performance measures (i.e., A') intended to reflect different levels of false recognition.

Table 1 presents three measures of recognition memory performance (A') for all three of our experiments. Table 2 presents the proportion of remember, know, and old judgments for each condition and type of

item for Experiment 1. Table 3 presents the proportion of remember, know, and old judgments for each condition and type of item for Experiment 2. One type of A' contrasts old responses to studied items vs. old responses to new items (*old–new A'* hereafter), which we utilize as our measure of accurate (true) recognition. The second type contrasts old responses to lure items from studied themes with old responses to new lures (*lure A'* hereafter). Lure A' can be thought of as a measure of how likely participants were to false alarm to a lure item when its corresponding theme was studied compared to when no items related to the lure were studied, with higher values of lure A' indicating higher levels of false recognition. The third type of A' was computed on study items vs. lure items (*old-lure A'* hereafter), which provides a measure of how well participants were able to discriminate between items that were presented at study and semantically related information that was not presented at study. We utilize lure A' and old-lure A' as our measures of false recognition. All analyses were performed with an alpha level of .05.

Table 1
Mean old–new A', lure A', and old-lure A' as a function of font condition for Experiments 1–3

	Font condition	Old–new A'	Lure A'	Old-lure A'
Experiment 1	Correlated	.940	.864	.656
	Uncorrelated	.926	.771	.767
	Unique	.922	.685	.810
Experiment 2	Correlated	.937	.832	.691
	Uncorrelated	.925	.747	.761
	Unique	.940	.714	.831
Experiment 3	Correlated	.931	.806	.707
	Standard	.918	.801	.669
	Unique	.933	.719	.817

Table 2
Mean proportion of remember, know, and old responses as a function of font condition and item type for Experiment 1

Font condition	Item type	Test response		
		Remember	Know	Old
Correlated	Old	.71	.18	.89
	Lure	.52	.25	.77
	New	.04	.06	.10
	New lure	.06	.11	.17
Uncorrelated	Old	.63	.22	.85
	Lure	.23	.25	.48
	New	.03	.07	.10
	New lure	.06	.09	.16
Unique	Old	.69	.20	.89
	Lure	.17	.30	.47
	New	.04	.11	.15
	New lure	.05	.18	.23

Note. Old responses are the summation of remember responses and know responses.

Table 3
Mean proportion of remember, know, and old responses as a function of font condition and item type for Experiment 2

Font condition	Item type	Test response		
		Remember	Know	Old
Correlated	Old	.66	.27	.93
	Lure	.41	.38	.78
	New	.04	.12	.15
	New lure	.07	.14	.21
Uncorrelated	Old	.53	.35	.88
	Lure	.16	.44	.60
	New	.03	.11	.15
	New lure	.05	.17	.21
Unique	Old	.65	.27	.92
	Lure	.09	.39	.48
	New	.02	.11	.13
	New lure	.02	.14	.16

Note. Old responses are the summation of remember responses and know responses.

Experiment 1 results

The primary result from Experiment 1 was that font condition had its strongest effects on false alarms to lure items (see Table 2). This result is evident in analyses of lure A', old-lure A', and the proportion of remember responses to lure items, which are presented below.

Participants in the three groups were approximately equivalent in terms of old–new A', $F(2, 87) = .94$, $MS_e = .003$, $\omega^2 < .01$. However, participants did perform differently on lure items across the three font conditions, $F(2, 87) = 23.39$, $MS_e = .01$, $\omega^2 = .33$. Tukey's *HSD* post hoc comparisons revealed that lure A' reliably differed between all three study conditions ($HSD = .063$). Thus, lure A' provides evidence that false recognition increased from the unique ($M = .68$) to the uncorrelated ($M = .77$) to the correlated conditions ($M = .86$). The influence of font condition on false recognition was also evident in analyses of old-lure A'. A one-way between-participants ANOVA revealed a significant difference among the font conditions, $F(2, 87) = 11.464$, $MS_e = .016$, $\omega^2 = .19$. Tukey's *HSD* post hoc comparisons ($HSD = .08$) revealed that old-lure A' was lower in the correlated condition ($M = .66$) than the unique condition ($M = .81$) and the uncorrelated condition ($M = .77$). However, the comparison between the unique and uncorrelated conditions did not reach the criterion for significance.

The difference in lure item performance across the font conditions was primarily reflected in the proportion of remember responses. Remember responses to lure items reliably differed across the font conditions ($F(2, 87) = 27.838$, $MS_e = .038$, $\omega^2 = .37$). Tukey's *HSD* post hoc comparisons ($HSD = .13$) indicated that there were more remember responses to lure items in the correlated condition ($M = .52$) than in the other two conditions ($M = .23$ and $M = .17$ for the uncorrelated

and unique conditions, respectively), which did not reliably differ.

Experiment 2 results

The results of Experiment 2 largely replicated the results of Experiment 1. In particular, the differences in false recognition observed as a function of font condition in Experiment 1 were again evident in the results of Experiment 2 (see Tables 1 and 3). In contrast to Experiment 1, participants performed differently in terms of old–new A' across the three font conditions, $F(2, 70) = 6.91$, $MS_e = .003$, $\omega^2 = .14$. Tukey's *HSD* post hoc comparisons ($HSD = .0104$) indicated that A' was greater in the correlated ($M = .937$) and unique ($M = .940$) conditions relative to the uncorrelated condition ($M = .925$), but that the correlated and unique conditions did not reliably differ. Lure A' and old-lure A' showed similar patterns of performance as a function of font condition as was found in Experiment 1. Specifically, performance differed in terms of lure A' across the three study conditions, $F(2, 70) = 10.00$, $MS_e = .013$, $\omega^2 = .20$. Tukey's *HSD* post hoc comparisons ($HSD = .066$) indicated that lure A' was higher in the correlated condition ($M = .832$) than in the unique ($M = .714$) and uncorrelated ($M = .747$) conditions, which did not reliably differ. Finally, the effects of font condition on lure items was also evident in old-lure A'. A one-way repeated measures ANOVA demonstrated a difference among the font conditions, $F(2, 70) = 33.62$, $MS_e = .005$, $\omega^2 = .47$. Tukey's *HSD* post hoc comparisons ($HSD = .041$) showed that all three groups reliably differed from one another, with A' decreasing from the unique condition ($M = .831$) to the uncorrelated condition ($M = .761$) to the correlated condition ($M = .691$).

Also replicating Experiment 1, the patterns of A' were primarily reflected in the patterns of remember

responding. Specifically, remember responses to old items differed reliably across the font conditions ($F(2, 70) = 22.287$, $MS_e = .0079$, $\omega^2 = .37$). Tukey's *HSD* post hoc comparisons ($HSD = .05$) demonstrated that the correlated ($M = .66$) and unique ($M = .65$) conditions showed higher rates of remember responding than the uncorrelated condition ($M = .53$), but that the correlated and unique conditions did not reliably differ from one another. Additionally, remember responses to lure items reliably differed across the font conditions ($F(2, 70) = 62.682$, $MS_e = .0158$, $\omega^2 = .49$). Tukey's *HSD* post hoc comparisons ($HSD = .07$) indicated that all three font conditions reliably differed from one another, with remember responding being greatest in the correlated condition ($M = .41$), least in the unique condition ($M = .09$), and intermediate in the uncorrelated condition ($M = .16$).

The critical finding in these two experiments was that the patterns of false recognition largely replicated across Experiments 1 and 2, despite the fact that font condition was manipulated between participants in Experiment 1 and within participants and lists in Experiment 2. The occurrence of differences in false recognition as a function of font condition in both experiments is inconsistent with the predictions of the distinctiveness heuristic hypothesis for the effects of font condition on false recognition. If a global distinctiveness heuristic were producing the differential suppression of false recognition in the font conditions, one would expect those differences to disappear when font condition is manipulated within participants and lists, a result that did not occur. The results of these two experiments are also problematic for other explanations of false recognition that rely on differences in decision strategies across experimental conditions (e.g., Hirshman, 1995; Miller & Wolford, 1999), because decision strategy shifts are unlikely to occur within a single test list (Stretch & Wixted, 1998). However, the results of Experiments 1 and 2 are consistent with memory-based explanations of false recognition, such as the item specific and relational processing account (Einstein & Hunt, 1980; Hunt & Einstein, 1981). Recall that memory based explanations would predict that the effects of font condition on false recognition should replicate across between and within participants designs, exactly the result obtained in these two experiments.

One final question concerns the relationship between the present results and the prior literature. While we have demonstrated that the characteristics of the visual features in which study items are presented affects false recognition, it is not clear how these results fit within the greater context of the false recognition literature. Specifically, the interpretation of performance in the unique and uncorrelated conditions depends on how one views the level of performance in the correlated condition. Thus, it is possible that the presentation of items in the correlated font condition leads to an increase or decrease

in false recognition relative to the manner in which items have been presented visually in previous experiments (e.g., Arndt & Hirshman, 1998; Gallo et al., 2001; Smith & Hunt, 1998). To investigate this question, we replaced the uncorrelated font condition with a standard font condition in Experiment 3. In the standard font condition, all study items and lures were presented in a common font (Helvetica).

Experiment 3

Method

Participants

Thirty-six students from Carnegie Mellon University participated in order to satisfy a research appreciation requirement.

Materials and design

This experiment again manipulated three font conditions. The correlated condition and the unique condition were the same as described in Experiment 2. However, the uncorrelated condition was replaced with a condition in which all study items were experienced in a single common font. The font chosen for display of these items was Helvetica. Four themes in each study list were presented in Helvetica, as well as a comparable number of new test items from four unstudied themes, and the lure items corresponding to those themes, both studied and unstudied.

Procedure

The procedure of Experiment 3 was identical to that of Experiment 2 with two exceptions. The uncorrelated condition was replaced with the standard condition. As a consequence, 105 different fonts were experienced by participants in this experiment, of which 104 were from the same pool described for Experiments 1 and 2, and one (Helvetica) was a common, non-distinctive looking font utilized to display study and test items assigned to the standard condition. As with Experiment 2, assignment of: (1) fonts to the correlated and unique conditions; (2) themes to experimental condition; (3) fonts to themes in the correlated and unique conditions (and items in the unique condition); (4) themes to be studied or unstudied; (5) themes to serial position within the study list; (6) items to serial position within a studied theme; and (7) items to serial position within the test list were all determined randomly for each participant using a computer algorithm.

Results and discussion

The results of Experiment 3 are presented in terms of old–new A', lure A', and old-lure A' in Table 1, and in

Table 4
Mean proportion of remember, know, and old responses as a function of font condition and item type for Experiment 3

Font condition	Item type	Test response		
		Remember	Know	Old
Standard	Old	.68	.19	.88
	Lure	.44	.28	.72
	New	.05	.10	.15
	New lure	.09	.14	.23
Correlated	Old	.73	.15	.89
	Lure	.47	.21	.67
	New	.03	.09	.12
	New lure	.04	.15	.18
Unique	Old	.69	.20	.89
	Lure	.11	.32	.44
	New	.02	.09	.11
	New lure	.04	.10	.14

Note. Old responses are the summation of remember responses and know responses.

terms of the proportion of remember, know, and old responses in Table 4. Across the three font conditions, participants performed differently in terms of old–new A', $F(2, 70) = 5.18$, $MS_e = .0005$, $\omega^2 = .10$. Tukey's *HSD* post hoc comparisons ($HSD = .0126$) showed that A' was greater in the unique condition ($M = .933$) and correlated conditions ($M = .931$) than in the standard condition ($M = .918$), but that the unique and correlated conditions did not reliably differ from one another. Lure A' also differed across the three study conditions, $F(2, 70) = 6.97$, $MS_e = .012$, $\omega^2 = .14$. Tukey's *HSD* post hoc comparisons ($HSD = .063$) showed that A' was lower in the unique condition ($M = .719$) than in the correlated ($M = .806$) and standard ($M = .801$) conditions, which did not reliably differ. The effects of font condition on false recognition were also evident in old-lure A'. A one-way repeated measures ANOVA demonstrated a difference among the font conditions, $F(2, 70) = 26.429$, $MS_e = .008$, $\omega^2 = .41$. Tukey's *HSD* post hoc comparisons ($HSD = .051$) showed that A' was greater in the unique condition ($M = .817$) than in the correlated ($M = .707$) and standard ($M = .669$) conditions, which did not reliably differ.

Finally, the effects noted above for A' were reflected in the patterns of remember responding for lure items, but not for old items. The proportion of remember responses to old items did not reliably differ across the font conditions ($F(2, 70) = 2.228$, $MS_e = .0138$, $\omega^2 = .03$), while the proportion of remember responses to lure items did ($F(2, 70) = 33.678$, $MS_e = .0408$, $\omega^2 = .47$). Tukey's *HSD* post hoc comparisons ($HSD = .11$) indicated that fewer remember responses were given to items in the unique condition ($M = .11$) than in the correlated ($M = .47$) and standard ($M = .44$) conditions, which did not reliably differ.

Thus, false recognition in the correlated condition was found to be roughly comparable in magnitude to false recognition in the standard visual presentation condition. Consequently, performance in the unique and uncorrelated conditions can be interpreted as producing suppression of false recognition relative to the modal case of visual presentation used in past studies (e.g., Arndt & Hirshman, 1998; Gallo et al., 2001; Smith & Hunt, 1998). This suppression is especially notable in the context of prior studies investigating the effect of study and test modality on false recognition (Gallo et al., 2001; Smith & Hunt, 1998). These studies demonstrated that the conditions utilized in the present experiment (visual study and visual test) produce lower levels of false recognition relative to conditions in which study lists, test lists, or both were presented in the auditory modality.

General discussion

As noted above, aspects of these data pose difficulty for an explanation based upon decision-based processes, such as the operation of a global distinctiveness heuristic (Schacter et al., 1999), and favor an account based upon memory processes independent of decision processes. Specifically, false recognition was found to vary across the font conditions in all three experiments reported here, even when font condition was manipulated within participants. The comparison of the effects of font condition within and between participants provides a critical test of the distinctiveness heuristic in particular and decision-based approaches to false recognition reduction in general. As detailed above, as well as by Schacter et al. (1999) and Dodson and Schacter (2001), a distinctiveness heuristic is unlikely to produce differential reduction of false recognition across conditions in a

within-participants design (see, e.g., Stretch & Wixted, 1998 for similar conclusions about within list decision strategy shifts). Thus, any differences in false recognition observed when font condition is varied within participants and lists is likely to reflect memory-based processes.

Although these data favor memory based accounts of false recognition over decision based accounts, they do not appear to be sufficient to discriminate among different memory based accounts. While we outline one such account that appears to be a promising explanation of the effect of the encoding of visual features on false recognition, we recognize that other explanations are plausible, a point that we expand upon in our concluding comments. The particular account that we outline here combines the encoding process elements from the distinction between item specific and relational processing at study (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993; Smith & Hunt, 1998) with the representational and retrieval mechanisms of an activation based memory model (e.g., Reder et al., 2000). The resultant account not only accommodates how true and false recognition are affected by font condition, but also provides an explanation of why the patterns of true and false recognition were generally reflected in participants' "remember" responses.

Recall that relational information arises from the processing of the commonalties among study items, while item-specific information arises from the processing of the differences between study items. The encoding of relational information, while typically beneficial to memory performance, would tend to not only improve the recognition of study items, but also increase the false recognition of semantically related lure items. In contrast, the encoding of item-specific information should only increase accurate recognition of study items. Thus, the study conditions that are conducive to the encoding of relational information will increase both levels of accurate recognition of study items and false recognition of lure items, while study conditions that are conducive to the encoding of item-specific information will only increase accurate recognition of study items. Finally, variants of this approach assume that the encoding of item specific and relational information compete for finite encoding resources (e.g., DeLosh & McDaniel, 1996; Serra & Nairne, 1993). Specifically, the extent to which study items promote attention to and the encoding of item-specific information determines how much relational information is encoded. Thus, conditions that are conducive to the encoding of item-specific information should diminish the encoding of relational information relative to conditions that do not promote the encoding of item-specific information.

The explanation of this theory rests on the assumption that item specific encoding is most prevalent in the unique condition, less prevalent in the uncorrelated condition, and least prevalent in the correlated condition. Because item specific and relational encoding compete for encoding resources, relational encoding would be least prevalent in the unique condition, more prevalent in the uncorrelated condition, and most prevalent in the correlated condition. Thus, in the unique condition, the unusual visual patterns inherent in the fonts would be expected to produce high levels of item-specific encoding, because such encoding benefits from the presence of factors that serve to distinguish one item from another such as non-overlapping visual information. In contrast, the correlated condition would not be expected to provide very salient item-specific information, because the study items shared both semantic and perceptual features, leaving relatively little information available to encode that would distinguish among items. Thus, performance in the correlated condition should be primarily driven by relational information. Finally, in the uncorrelated condition, the salience of item-specific information falls somewhere between the unique condition and the correlated condition. Because fonts in this condition were repeated as often as in the correlated condition, there would be less of an expectation of item-specific encoding relative to the unique condition. However, because items in each of the themes were presented in several different fonts, participants would still have a stronger basis for encoding item-specific information relative to the correlated condition (i.e., the perceptual characteristics associated with different study items varied within themes).

Recollective experience and false recognition

The seminal paper of Roediger and McDermott (1995), in addition to reporting high levels of false recall and false alarms to lure items also reported that participants often claimed to "remember" lure items. This recollective report is important because it suggests that participants believe that the lure item was presented and that the false recognition of lure items is not due to a bias to accept a lure because it is related to numerous study items (Miller & Wolford, 1999; Roediger & McDermott, 1999). In the present experiments, the font manipulations had their primary influence on participants' tendency to give a remember response, indicating that participants' belief that the lure items were presented was influenced by the font manipulation. Here, we combine elements of the item specific and relational processing account with an extant model of representation and retrieval processes in an effort to explain the differential recollection phenomenology for lure items as a function of font condition.

A mechanistic account of recognition, proposed by Reder et al. (2000), claims that when a word is presented for study, it activates its corresponding *concept node* in memory (Reder et al., 2000, 2002b). Further, during a

study event, a trace is laid down that links the activated concept of the word to the study context. The trace is represented as an *episode node* that is associated with both the concept node and an experimental *context node*. In the recognition memory experiments of Reder et al. (2000, 2002b), the study lists were comprised of unrelated words, and thus there was no need to consider the role that relational or organizational representations may play in recognition memory. However, using the same type of network representation and spread of activation processes, Reder and Ross (1983; see also Reder & Anderson, 1980) posited that when subjects encounter thematically related material, such as multiple sentences related to the theme of washing clothes, subjects create a *theme node*. It seems reasonable to assume that in situations where the relationships among study items are salient, such representations are constructed in memory, and are associated with the concept nodes of the related study items. Further, as occurs for other representations that are constructed in a particular context, the theme node is associated with the experimental context. The final assumption that is required to explain recollective

reports for lure items is that subjects often *implicitly generate* (Anaki, Faran, & Avishai, 2001; McDermott, 1997; Underwood, 1965) the thematically related target word during the study of its associates, causing its concept node to be associated with its appropriate theme node. Fig. 2 provides a graphical depiction of the representational scheme outlined above.

Retrieval from this network representation is governed by the same spreading activation processes proposed by Reder et al. (2000). At test, the concept node associated with the probe and the experimental context node are activated and activation spreads across all of their associations. Positive recognition memory decisions can be made based upon one of two sources of information: the activation level of memory traces associated with the experimental context or the activity level of an item's conceptual representation. When the activation level of a memory trace associated with the experimental context (i.e., a thematic or episodic representation) is sufficiently high, participants will respond based upon recollection (i.e., they will give it a *remember* response). However, when the activation of the theme or
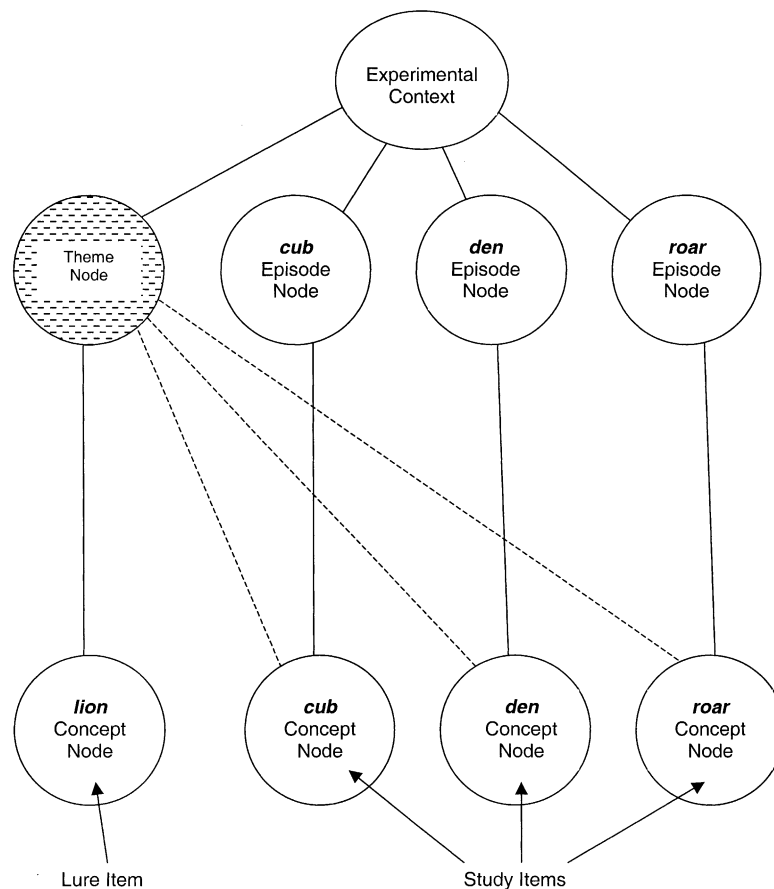

Fig. 2. Semantic network representation of a studied theme.

episode node is not sufficient to produce recollection, participants may still respond based upon the activation level of the concept node. In contrast to responses based upon activation of a thematic or episodic representation, the activation level of conceptual representations, because they are not directly associated with the experimental context, provide participants with a sense of how familiar an item is. Thus, if the conceptual representation (i.e., the concept node) is sufficiently active, participants may still endorse an item as old, but they will do so based upon a feeling of familiarity (i.e., they will give the item a *know* response) rather than recollection.

The assumptions of differential encoding (item specific vs. relational) in the font conditions can be combined with the representation and retrieval assumptions outlined above, producing the differences in recollective reports across conditions observed in these experiments for old items and lure items. Specifically, item-specific processing would tend to enhance the strength of episode nodes, while relational processing would tend to enhance the strength of the theme nodes. Thus, for example, in the unique font condition, episode nodes would tend to be strengthened more than in the other two conditions, while theme nodes would be strengthened the least. In the correlated condition, theme nodes would tend to be strengthened more than in the other two conditions, while episode nodes would tend to be strengthened the least.

Given that (1) the strength of the thematic and episodic representations are proposed to affect remember judgments; and (2) the strengthening of thematic and episodic representations is proposed to vary across conditions in accord with the principles of item specific and relational processing previously outlined, this explanation expects the patterns of remember responding to be similar to those of recognition performance as measured by sensitivity measures such as A'. Specifically, study items in the correlated condition tend to be judged as recollected based upon the high level of activation of the thematic representation, while study items in the unique condition tend to be judged as recollected primarily based upon the high level of activation of their individual episodic traces. However, lure items only have one possible source for their positive recollection: the activation level of the thematic representation. Because this representation will be more active in the correlated condition, participants will falsely recollect lure items more often than in the unique and uncorrelated conditions. Further, because this representation is weakest in the unique condition, participants will falsely recollect its presentation less often than in the correlated and uncorrelated conditions. Thus, the network representation of Reder et al. (2000; see also Reder & Ross, 1983), combined with the differential strengthening of thematic and episodic representations according to the

principles of item specific and relational processing provides an account of not only true and false recognition across conditions (measured in terms of A'), but also of the fact that these patterns were primarily reflected in participants' remember responses.

Concluding comments

While we have elaborated one theoretical position's account of these data, we do not intend to imply that other theoretical explanations of these data are implausible. Rather, it seems likely that other perspectives that propose a similar dichotomy to that between item and relational information could also explain these data (e.g., Brainerd, Reyna, & Poole, 2000; Humphreys, 1978; Jacoby, 1991; Jones & Jacoby, 2001; Rajaram, 1996, 1999; Yonelinas, 1997). To take one example, fuzzy trace theory (Brainerd et al., 2000) would offer a comparable explanation of these results. In particular, fuzzy trace theory proposes that recognition memory is due to the encoding of two different types of information, verbatim information and gist information. Thus, verbatim information, the encoding of the physical features of study items, roughly maps on to item-specific information, while gist information, the encoding of shared meanings among study items, roughly maps on to relational information. Provided the assumption of trade-offs in encoding of gist and verbatim information that is a characteristic of the item-specific and relational information explanation, fuzzy trace theory would explain these results in a manner similar to that of the item-specific and relational processing account.

In summary, we have demonstrated that differences in the perceptual characteristics of study items can influence both true and false recognition. The patterns of false recognition in our experiment are inconsistent with one account of false recognition suppression following the study of distinctive visual information, the distinctiveness heuristic (Schacter et al., 1999). At a more general level, these results are incompatible with the most straightforward conceptualizations of the influence of decision based processes on false recognition. Rather, these data favor a memory based explanation of false recognition reduction.

Acknowledgments

References

Anaki, D., Faran, Y., & Avishai, H. (2001). The false memory and mirror effects: The role of familiarity and backward association in creating false recollection. Poster presented at the 42nd Meeting of the Psychonomic Society, Orlando, FL.

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language, 39*, 371–391, doi:10.1006/jmla.1998.2581.

Brainerd, C. J., Reyna, V. F., & Poole, D. A. (2000). Fuzzy-trace theory and false memory: Memory theory in the courtroom. In D. F. Bjorklund (Ed.), *False-memory creation in children and adults: Theory, research, and implications* (pp. 93–127). Mahwah, NJ: Erlbaum.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1136–1146.

Diana, R., Peterson, M., & Reder, L.M. (2002). The role of spurious feature familiarity in recognition memory. Manuscript under review.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*, 155–161.

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46*, 782–803, doi: 10.1006/jmla.2001.2882.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 588–598.

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 339–353.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 302–313.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1306–1323.

Humphreys, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning and Verbal Behavior, 17*, 175–187.

Hunt, R. R., & Einstein, G. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421–445.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4*, 577–581.

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*, 82–102.

McDermott, K. B. (1997). Priming on perceptual implicit tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review, 4*, 582–586.

McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What is the connection between true and false memories. The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1177–1194.

Miller, M. B., & Wolford, G. L. (1999). Theoretical Commentary: The role of criterion shift in false memory. *Psychological Review, 106*, 398–405.

Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Available: http://www.usf.edu/Free-Association/.

Payne, D. M., Elie, C. L., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*, 261–285, doi:10.1006/jmla.1996.0015.

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 365–377.

Rajaram, S. (1999). Assessing the nature of retrieval experience: Advances and challenges. In B. H. Challis & B. M. Velichovsky (Eds.), *Stratification in cognition and consciousness* (pp. 255–275). Amsterdam: John Benjamins.

Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A., & Ayers, M. A. (2002a). A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. *Journal of Experimental Pyschology: Learning, Memory, and Cognition, 28*, 138–152.

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

Reder, L. M., Donavos, D., & Erickson, M. A. (2002b). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition, 30*, 312–323.

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

Reder, L. M., & Ross, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 55–72.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Roediger, H. L., & McDermott, K. B. (1999). False alarms about false memories. *Psychological Review, 106*, 406–410.

Russell, W.A., & Jenkins, J.J. (1954). The complete Minnesota norms for responses to 100 words from the Kent–Rosanoff

word association test. (Tech. Rep. No. 11, Contract N8 ONR 66216, Office of Naval Research). Minneapolis: University of Minnesota.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24, doi: 10.1006/jmla.1998.2611.

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition, 21*, 34–40.

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

Smith, R., & Hunt, R. R. (1998). Presentation modality affects false recognition. *Psychonomic Bulletin & Review, 5*, 710–715.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*, 494–500.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 681–690.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26*, 1–12.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122–129.

Wickens, T. D., & Hirshman, E. (2000). False memories and statistical decision theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review, 107*, 377–383.

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107*, 368–376.

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