Carnegie Mellon University
Research Showcase

Department of Psychology | Dietrich College of Humanities and Social Sciences

10-1-2005

# The list strength effect: A contextual competition account

R A. Diana

L M. Reder
cherdema@andrew.cmu.edu

Follow this and additional works at: http://repository.cmu.edu/psychology

Recommended Citation
Diana, R A. and Reder, L M., "The list strength effect: A contextual competition account" (2005). *Department of Psychology.* Paper 373.
http://repository.cmu.edu/psychology/373

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase. For more information, please contact research-showcase@andrew.cmu.edu.

*Memory & Cognition*
*2005, 33 (7), 1289-1302*

# The list strength effect: A contextual competition account

RACHEL A. DIANA and LYNNE M. REDER
*Carnegie Mellon University, Pittsburgh, Pennsylvania*

Research on the list strength effect (LSE) has shown that learning some words on a list more strongly than others impairs memory for the weakly learned words when tested with a recall task. Norman (2002) demonstrated that the LSE also occurs within the recollection process of a recognition test. In this study, a mechanistic dual-process account of the LSE was tested that simultaneously makes predictions concerning additional sources of context in interference effects. In two experiments, we attempted to replicate Norman's (2002) findings and provide the basis for our modeling efforts. We found evidence for a recollection LSE in raw measures of responding, with memory performance also benefiting from reinstatement of perceptual characteristics at test. However, large differences in the hits between the lists were accompanied by small differences in false alarms, such that when $d'$ is calculated, differences between the lists are not significant. We propose an account of the LSE whereby the actual effect of competition between items on the list is small, although present, and difficult to distinguish from large effects of bias due to the strength manipulations. We argue that our findings provide support for a mechanistic explanation of LSE that is based on competition of source activation and changes in the thresholds for responses.

The study of interference is central to understanding how memory works. Indeed, recent attention to the mechanisms underlying interference has allowed better understanding of memory processes in general (e.g., Anderson, 1981; Chandler, 1991; Kim & Glanzer, 1995; Murnane & Shiffrin, 1991a; Willis & Underwood, 1983; Yonelinas & Jacoby, 1994). In particular, the list length and list strength paradigms have been popular methods for investigating interference in recall and recognition tasks. List length manipulations test the sensitivity of memory to the number of items on a list, whereas list strength manipulations test memory sensitivity to the relative strength of items on a list. These types of experiments can help determine how information is related in memory and how information that is stored in the same context influences memory for other information.

List length experiments manipulate the number of items on a list and compare memory for lists with different numbers of items. List length manipulations have been found to affect both recognition and recall (e.g., Gillund & Shiffrin, 1984; Gronlund & Elam, 1994; Ohrt & Gronlund, 1999). Some researchers have questioned whether previously demonstrated list length effects (LLEs) in recognition were due to methodological confounds, such as retention interval, rather than to interference (Dennis & Humphreys, 2001; Murdock & Kahana, 1993). However, the results have held under carefully controlled conditions (Cary & Reder, 2003; Ohrt & Gronlund, 1999). Thus, at least under some conditions, increasing the number of items on a list makes it more difficult to remember any one of the items.

Tulving and Hastie (1972) were the first to investigate the effects of stronger learning of some items on memory for more weakly learned items. They found a list strength effect (LSE) for free recall, such that when some words on a list were presented twice, performance for the singly presented items was impaired. Ratcliff, Clark, and Shiffrin (1990) found some evidence for an LSE in the cued recall paradigm, but they were unable to find a consistent LSE for recognition. Many other researchers have also found little or no evidence for impairment in recognition tasks based on list strength (e.g., Hirshman, 1995; Murnane & Shiffrin, 1991a; Ratcliff, Sheu, & Gronlund, 1992; Yonelinas, Hockley, & Murdock, 1992). Murnane and Shiffrin (1991b) were the first to show consistent evidence of an LSE for recognition, but only when each repetition of a word was presented within a unique sentence. They proposed that each presentation in a unique context is stored as a separate memory trace, creating a list length manipulation, rather than a list strength manipulation in which an individual trace would be strengthened. The evidence in the literature is that an LSE does not occur for simple recognition unless the context is varied with each repetition.

Researchers have also examined how differences in item strength on mixed lists and pure lists affect participants' biases to respond *old* or *new*. Meta-analyses of null LSE findings have shown that bias is more conservative

This work was supported by Grant 2-R01-MH52808 from the National Institute of Mental Health to L.M.R. R.A.D. was supported by an NSF graduate fellowship and, in part, by the Center for the Neural Basis of Cognition at Carnegie Mellon University and the University of Pittsburgh. We thank Heekyeong Park for helpful comments on an earlier version of the manuscript. Correspondence concerning this article should be addressed to R. A. Diana, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213 (e-mail: rdiana@andrew.cmu.edu).

 Copyright 2005 Psychonomic Society, Inc.

for mixed weak items than for pure weak items (Chappell & Humphreys, 1994; Hirshman, 1995). Hirshman also showed that pure strong items produce a more conservative bias than do mixed strong items. This bias seems to be present in most, if not all, LSE experiments. Rather than producing a mirror effect, such that hits increase and false alarms decrease, the results tend to show an increase in both hits and false alarms. The resulting measures of $d'$ do not show significant effects of list strength.

Dual-process theories of recognition memory propose that items are recognized on the basis of either a familiarity process or a recollection process (e.g., Jacoby, 1991; Reder et al., 2000; Tulving, 1985; Yonelinas, 1994). The null LSE findings were difficult for dual-process theories to explain, because these theories propose a recollection process in recognition that is similar to that involved in recall. Therefore, if recall is affected by a list strength manipulation, the recollection process in recognition should also be affected. On the basis of this logic, Norman (1999, 2002) tested the LSE for *remember*/*know* (Tulving, 1985) responses. *Remember* responses are believed to access the recollection process in recognition, and thus any effect of list strength should occur in these responses. Norman explained the previous null LSE findings as the result of compensation from the familiarity process when the recollection process is impaired. In fact, Norman did find that the LSE occurred in *remember* responses, but not in overall responding, thus replicating the previous null LSE for recognition, but also demonstrating impairment of recollection, due to interference.

Our goal is to provide and test a mechanistic account of the recollection LSE within the source of activation confusion (SAC) model (Reder et al., 2000), a well-specified dual-process account that could explain Norman's (1999, 2002) data. In addition to proposing a set of mechanisms behind recognition processing, SAC makes quantitative predictions concerning the contributions of recollection and familiarity to recognition. We also use the SAC model to develop new predictions regarding memory performance when both list strength and context are manipulated. The SAC explanation of the recollection LSE is based on contextual factors, and therefore, a manipulation of contextual variables provides an interesting test of this account.

As is illustrated in Figure 1, SAC is a model that represents concepts in a localist network. Activation spreads among concepts in accordance with the relative strength of a connection (as compared with the strength of the competitors) and the amount of activation a concept has available to spread. During encoding, memory traces representing the information seen at study are created and bound to the experimental/list context. The trace that binds a word to its experimental context is called an *episode node*. When a probe word is presented on a recognition test, the more active the episode node, the more likely a participant is to recollect seeing the word in the experiment. When the node representing the word itself is more active, participants have a stronger feeling of familiarity toward that item. Nodes and links between nodes are strengthened with each presentation. Specific equations used to calculate activations and responses in SAC are described in Reder et al.'s (2000) article.

The nature of the spreading activation (according to the number and strength of the links) is key to the SAC explanation of interference. SAC explains the LLE by proposing that increasing the number of items on the list decreases the amount of activation that spreads from the experimental/list context to any episode node. Thus, the ability to recollect the items is reduced. As SAC would predict, Cary and Reder (2003) found that the decrease in performance with longer lists is manifested as a decrease in *remember* responses. SAC makes similar predictions regarding the effect of interference on recollection for the LSE.

The list strength paradigm involves a comparison between three types of lists. The pure weak list is made up of items that are learned to a weak standard (usually one presentation per item). The pure strong list has the same number of items, but each item is learned to a higher standard (usually multiple repetitions of each item or lengthened presentation times). The mixed list is composed of half weak items and half strong items but contains the same number of unique words as do both pure lists. The key comparisons are between the weak items on the pure list

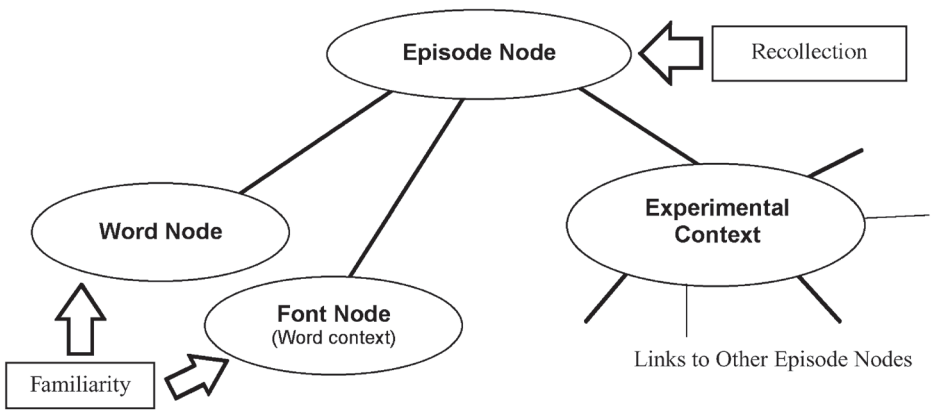


**Figure 1. A schematic representation of memory traces according to the SAC model.**

and the weak items on the mixed list, as well as between the strong items on the pure list and the strong items on the mixed list. If there is interference between the items on the list, weak items should be remembered better with the pure list than with the mixed list, due to the addition of the strong items on the mixed list. On the other hand, this interference should cause strong items to be remembered better with the mixed list than with the pure list. Strong items on the pure list must compete with a number of other strong items.

SAC predicts the LSE because increasing the strength of some items on the list decreases the amount of context activation that is available to other items on the list. The experimental context node is connected to all episode nodes on a list. Strong items on a list have stronger connections to the experimental context and, thus, receive more of the activation that spreads from the context node than do the weak items. In the pure weak list, all of the connections from episode nodes to the experimental context node are of the same strength.[1] At test, when the experimental context is activated, less activation spreads to the memory traces of the weak words (noted as episode nodes in Figure 1) on the mixed list than to the weak words on the pure list. This results in fewer recollection-based responses with the mixed list. However, the concept nodes for weak words should have the same resting level of activation on both lists. Obviously, strong words have higher familiarity than do weak words overall, but the familiarity does not differ depending on the list in which they are studied. It should be noted that the proportion of familiarity responses is inversely affected by the likelihood of recollection.

The account above explains why SAC predicts more *remember* responses to weak items on a pure weak list than to weak items on a mixed list; however, it does not explain why overall recognition is generally not affected by list strength manipulations. In SAC, recollection and familiarity are interdependent. The probability of a familiarity response is the probability of the concept node being above threshold, when no recollection response is made. The model states that participants make a familiarity response only if they are not able to recollect the item. Therefore, when ability to recollect decreases, the number of familiarity responses will increase if the activation of the concept node remains constant. This is the case in the LSE; familiarity of the words is unaffected by mixed list strength, but recollection decreases. Familiarity responses increase as recollection responses decrease. Experiments that measure responses as *old*/*new* only would be expected to see similar levels of responding for both pure weak and mixed list items.

The SAC account of the recollection LSE is based on differences in contextual factors between pure and mixed lists. Therefore, we were interested in the model's predictions regarding the effects of an additional and unique source of context that is independent of the list context. When each word has a unique context at study (e.g., an atypical font), reinstating that context at test should increase the number of *remember* responses, as compared with a condition in which the context (font) is different at test (see Reder, Donavos, & Erickson, 2002). That is, when the font matches and provides an extra source of activation at test, the effects of list strength can be overcome. These font match effects are seen in the *remember* responses, because when the font is reinstated, the font activation spreads to the episode node (consult Figure 1).

SAC predicts that the additional activation from the font source to the episode node will accrue along with the activation from the list context node, resulting in an increase in *remember* hits. Thus, the pure-list–font-match items are predicted to have the most *remember* responses, and the mixed-list–font-mismatch items are predicted to have the fewest *remember* responses. Because SAC makes these specific predictions regarding the LSE and additional sources of context (unique to each word), we are able to rigorously test our account of these effects.

## EXPERIMENT 1

### Method

**Participants**. The participants were 61 Carnegie Mellon undergraduates enrolled in psychology courses, who participated as partial fulfillment of a course requirement. Each participant was randomly assigned to the pure list (30 participants) or the mixed list (31 participants) condition.

**Design and Materials**. The design included two factors: list strength (between subjects) and font condition (within subjects). Pure list participants saw all the words on the list only once. Mixed list participants saw half of the words once and the other half six times. The font factor consisted of each word being presented at study with a unique font, and then half (in the match condition) being presented with the same font at test and the other half (in the mismatch condition) being presented with a font that had previously been seen with a different word. When words were repeated, they were presented in the same font for each repetition. All of the lures were seen in fonts that had been presented at study with different words. A number of dependent measures were examined: hits, false alarms, *remember* responses, *familiar* responses, $d'$, and reaction time (RT).

A total of 300 words were collected from the MRC Psycholinguistic Database at the University of Western Australia. The words had an average of 5.4 letters and average concreteness, familiarity, and Kučera–Francis frequencies of 583, 503, and 15.94, respectively. List construction was based on Norman's (2002) methods. Each study list consisted of 100 items randomly selected from the 300 total words for each participant. Half of these words were randomly assigned to be target items (always presented once), and the other 50 were interference items (presented once on the pure list and six times on the mixed list). Another 50 words were randomly selected as lure items to be shown along with the original 50 target items at test. The 200 fonts were collected from various Internet sources and were selected to be both distinctive and legible. Examples of some of the words and fonts used are shown in Figure 2. Fonts were randomly assigned to words for each participant, ensuring that any stimulus effects would not be systematic and would be treated as part of the random effects associated with subject variability. Repeated words were seen with the same font at each presentation during study.

**Procedure**. The participants were tested individually in sessions that lasted approximately 30 min. The experiment was run on a Macintosh computer using the PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). The session consisted of three phases: study phase, delay task, and test phase. The participants were told at the beginning that they would be asked to make judgments about words but were not specifically told that a memory test would be performed. At study, words were presented one at a

**Figure 2. Examples of words and fonts in Experiments 1 and 2.**

time for 1,500 msec. During this time, the participants were required to respond as to whether the font was appropriate for the meaning of the word. If the participants did not respond within 1,500 msec, a warning tone was heard. For pure lists, each target and interference item was presented once in the study phase. For mixed lists, each target and interference item was presented once, and then the interference items were presented five more times. Each cycle of interference items was completed before a second presentation of any one interference item occurred.

The difference in the number of presentations resulted in a longer study phase for the mixed list than for the pure list. The time between study and test was equalized by using a delay period and a distractor task. Immediately after completion of the study phase, timing of the delay period began. The delay period was 5 min long in the mixed list condition and 13.5 min long in the pure list condition. Thus, the time between study of target items and test on those items was equalized for the two groups. At the beginning of the delay, the participants were given instructions regarding the test phase. They were given standard *remember*/*know* instructions (Tulving, 1985), except that we replaced the term *know* with the term *familiar*, believing it to be less confusing for the participants. The participants were given an example of the difference between the two types of responses and were asked to give their own example to the experimenter to ensure understanding. The participants were also told to base their responses in the test phase on the word alone, regardless of the font. They were then asked to play Ultris (a freeware version of Tetris) for the remaining time of the delay period. All the participants played Ultris for at least 1 min.

During the test phase, a button box was used to ensure accurate RT collection. The participants were asked to be as accurate as possible, while still responding as quickly as possible. The button box had three keys, labeled *new*, *remember*, and *familiar*. The participants were instructed to use their middle fingers on the *new* key and their index fingers on the *remember* and *familiar* keys. Each word appeared on the screen, and the participants had an unlimited amount of time to respond. This procedure continued until all 50 targets (singly presented items) and 50 lure items were tested (presentation order of the items at test was randomly determined for each participant). No strong items were tested in the mixed list condition. We were concerned that the strong items would be near ceiling and, thus, would not add to our analyses.

## Results and Discussion

We had considered making all factors within subjects and, thus, tested early participants on both lists (one mixed and one pure weak). However, during the second study phase, the participants were aware that judging the font of each word was a distractor task and that a memory test would occur later. Neither of these pieces of information had been known during the first study phase, and they might have affected the way the participants performed the task. During a preliminary check of the data with 39 participants, we found that the participants' performance on the second list they received was poor (mean difference between the first and the second lists, $d' = .27$). When compared with performance on other lists, second-list performance was more similar to that for other second lists than to that for lists of the same type (pure or mixed). Given the reduction in performance and the difference in performance pattern, we treated the experiment as a between-subjects design and (for those participants that completed two lists) analyzed performance on the first list only. Later participants were only given one list.

The participants' responses were analyzed according to the percentage of hits and false alarms that were classified as *remember* responses, $d'$ values calculated with both *remember* responses and *old* responses overall, and RTs. Each dependent variable was analyzed separately for *remember* responses, for which we expected to find significant effects of list strength and font matching. We also analyzed *familiar* and overall *old* responses, for which we did not expect to find significant effects.

An ANOVA performed on the *remember* hit rates (see Figure 3) showed significant main effects for list type [$F(1,59) = 23.54$, $p < .001$] and font match condition [$F(1,59) = 140.85$, $p < .001$] such that there were more *remember* hits to words that did not have strong competitors on the list and words that were re-presented in the same font at test. These data support our predictions that words on the pure list would be better recollected, showing a significant LSE, and that words testing the match condition would be better recollected. There was no significant interaction between the variables.

We also analyzed false alarm responses, when judged both as *remember* and as *familiar* responses. As can be seen in Figure 3, overall, there were more *familiar* false alarms than *remember* false alarms and more false alarms by pure list participants than by mixed list participants. We performed $t$ tests separately for each response type, comparing pure list with mixed list false alarms. We found a significant effect for *remember* false alarms [$t(59) = 4.17$, $p < .001$] and *familiar* false alarms [$t(59) = 5.61$, $p < .001$], such that in both cases there were more false alarms to items on the pure list than to items on the mixed list.

We compared $d'$ when calculated with *remember* responses only with $d'$ when calculated collapsing across all *old* responses, as can be seen in Figure 4. The data show a significant effect of font match condition for both *remember* $d'$ [$F(1,59) = 133.75$, $p < .001$] and *old* $d'$ [$F(1,59) = 151.40$, $p < .001$]. We also found no differences based on list type; $d'$ for *old* responses were slightly better for the mixed list than for the pure list condition, but not reliably so ($p > .05$). When *remember* responses alone were examined, mean $d'$ was higher for the pure list in the font match condition than for the mixed list in the font match condition (see Figure 4), but the two scores were approximately equal in the font mismatch condition.

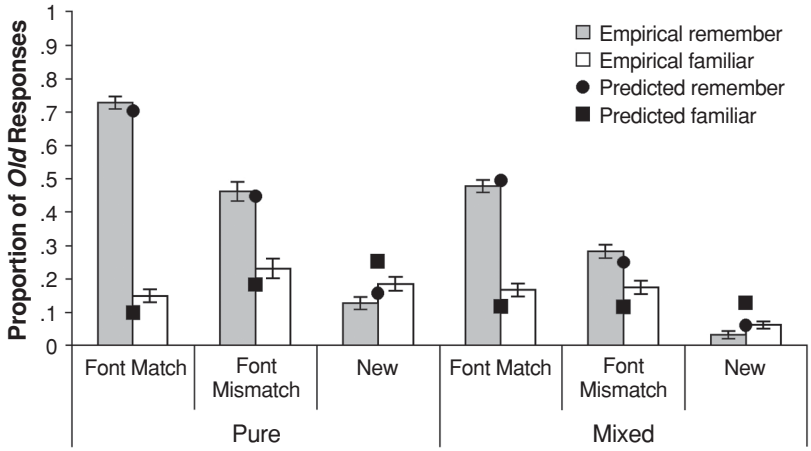


**Figure 3. Proportion of *remember* and *familiar* hits and false alarms as a function of match condition and list type in Experiment 1. Standard error bars are also plotted, along with the predicted data from the SAC simulation.**

Therefore, there was no main effect of list type for *remember* $d'$ ($p > .05$). It seems that the lack of effects are due to the larger false alarm rates for the pure list than for the mixed list, overshadowing the differences in the hit rates.

RTs may provide a more sensitive measure of the participants' memory abilities, and we examined them for *remember* and *familiar* responses (see Table 1). For *remember* responses, RTs were shorter for words when the font at test matched the font used during study [$F(1,55) = 17.81$, $p < .001$]. Interestingly, a significant main effect showed that people responded more quickly to words that were studied on the pure list than to those studied on the mixed list [$F(1,55) = 4.92$, $p < .05$]. There was no significant interaction between these two factors. The *familiar* responses were slower, overall, than the *remember* responses, as SAC would predict.[2] There were no significant differences for *familiarity* RTs for the font match condition or list type, nor was there an interaction. Therefore, as we would predict, the differences in RTs occurred in the *remember* responses, rather than in the *familiar* responses.

The evidence for a recollection LSE from this experiment occurred in RTs and in the *remember* hits, with false alarm differences between the pure and the mixed lists affecting $d'$ values that did not differ significantly. One final analysis was conducted in order to give more insight into the influence of bias versus recognition accuracy. The *familiar* hits should be an indicator of bias to some degree, because *familiar* responses are more likely to include guess responses or less confident *old* responses that are made due to bias. The number of *familiar* hits would be expected to increase along with *remember* hits in the pure list condition if the difference between pure and mixed lists was due entirely to a lower threshold. However, a two-way mixed ANOVA with factors of font match and list type did not show a significant effect of list type [$F(1,56) = 1.47$, $p = .23$]. Both the effects of font match [$F(1,56) = 7.83$, $p < .01$] and the interaction of font match and list type [$F(1,56) = 5.62$, $p < .05$] were significant.

This experiment supports SAC's mechanistic account of the LSE. As was predicted, additional sources of activation from a matching font contributed to the tendency

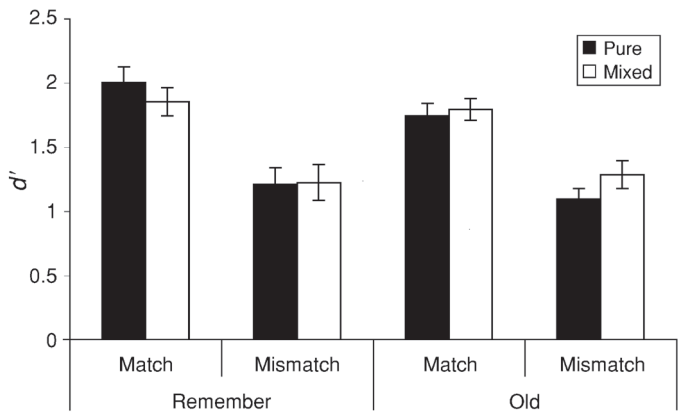


**Figure 4. Mean $d'$ calculated with *remember* responses and *old* responses as a function of match condition and list type in Experiment 1.**

**Table 1**
**Mean Reaction Times (in Milliseconds) for Experiment 1 as a Function of Response Type, List Type, and Font Match**

| | Font Match | | Font Mismatch | |
|---|---|---|---|---|
| Response | Pure List | Mixed List | Pure List | Mixed List |
| Remember | 1,264 | 1,600 | 1,497 | 1,719 |
| Familiar | 2,180 | 2,101 | 2,097 | 1,991 |

to recollect. This experiment also provides some evidence that interference occurs on the basis of list type, including a demonstration that competition from stronger list members reduces the speed of responding to the weaker list elements, as well as reducing the hit rate. We did expect to find a difference in $d'$ between the mixed and the pure lists, but this did not occur. However, while the *remember* hit rate was greater on the pure than on the mixed list, the *familiar* hit rate did not differ.

One reason that we might expect to have more difficulty finding a recollection LSE than did Norman (1999, 2002) is that we did not specifically attempt to increase participants' tendency to recollect. Norman's experiments were designed to determine whether an LSE would occur under optimum conditions for recollection, by maximizing feature overlap within lists, minimizing feature overlap between lists, avoiding floor and ceiling effects on individual items (Norman, 1999), and using plurality-reversed lures at test (Norman, 2002). Our experiments did not include these attempts to maximize recollection, and thus, our effect may be weaker. Our experiments are a test of the possibility that interference occurs naturally within lists.

## EXPERIMENT 2

In order to fully examine the LSE manipulation, Experiment 2 was designed to allow us to test the strong items on mixed and pure lists, as well as to strengthen our manipulation by making it within subjects and increasing the number of repetitions of strong items. SAC predicts that memory performance for strong items will be better on the mixed list than on the pure list. This is because each strong item must compete with more strong items for the activation from the experiment context in the pure list condition, whereas some of those strong items are replaced with weak items in the mixed list condition, thus reducing competition. In this experiment, we used shorter lists in order to reduce the effects of fatigue and between-lists interference that had forced us to perform only between-subjects analyses in Experiment 1. Also, the three lists were spread across two sessions, separated by 1 week, so that the participants saw no more than two lists in a day. These alterations allowed us to test all three types of lists with all participants. Finally, we increased the number of presentations of the strong items, which should cause a stronger effect of interference.

The same font manipulation was used in this experiment, where the fonts of the studied words were either the same as those at test or swapped with the fonts of other words. The font manipulation was useful in this experiment because it enabled us to examine the strong items in two different conditions. The font-match/strong-item condition should produce excellent memory performance, due to the extra cue for recollection, and thus it may be at or near ceiling. The font-mismatch/strong-item condition should be more difficult, thus allowing us to look at strong items that are not at ceiling.

### Method

**Participants**. The participants were 33 Carnegie Mellon students who participated as partial fulfillment of a course requirement, as well as receiving $7.

**Design and Materials**. The design was the same as that in Experiment 1, except that all lure items were presented in novel fonts. Also, all the participants were tested with all three lists. Each participant was randomly assigned a list order, such that all possible list orders occurred with approximately the same frequency. Once again, we were interested in a variety of dependent measures: hits, false alarms, *remember* responses, *familiar* responses, $d'$, and RT.

**Procedure**. Each study phase consisted of the presentation of 20 words, each in a unique font. The same pool of words and fonts was used as that in Experiment 1. On the pure weak list, all items were shown once. On the pure strong list, all items were repeated 11 times, in a random order, with each set of repetitions being completed before any of the words were shown again. On the mixed list, 10 items were shown once, whereas 10 items were repeated 11 times. All the list items were shown once before the strong items were repeated. The delay periods between study and test were altered, during which the participants played the Ultris video game, such that the time between the presentation of the first item on the study list and the first item on the test list was equal for all three lists.

Upon arrival, the participants were given *remember*/*know* instructions in both a written and an oral form, once again using the term *familiar* instead of *know*. They were required to give their own example of the difference between a *remember* and a *familiar* response before continuing with the experiment. Then the first study list was presented, followed by a delay and the test. In between the first list and the second list, the participants completed a separate experiment in which they were required to view photographs of people. The second study list was then presented, followed by a delay and the test. The participants returned 1 week later and completed the second half of the photograph experiment before seeing their third study list, followed by a delay and the test.

Other than these differences, the procedure was the same as that in Experiment 1.

### Results and Discussion

We analyzed the percentage of hits and false alarms that were classified as *remember* responses, $d'$ values calculated with both *remember* responses and *old* responses overall, and RTs. For hit rates, we performed three-way repeated measures ANOVAs with the factors of list type, item strength, and font match. It should be noted that *old* responses in the strong item, font match condition approached 100%. The ANOVA analyzing *remember* hit rates (see Figure 5A) showed significant main effects for item strength [$F(1,32) = 51.60, p < .001$], such that there were more *remember* hits to strong items than to weak items, and for font match condition [$F(1,32) = 63.38$, $p < .001$], such that words that were re-presented in the same font at test produced more *remember* hits than did

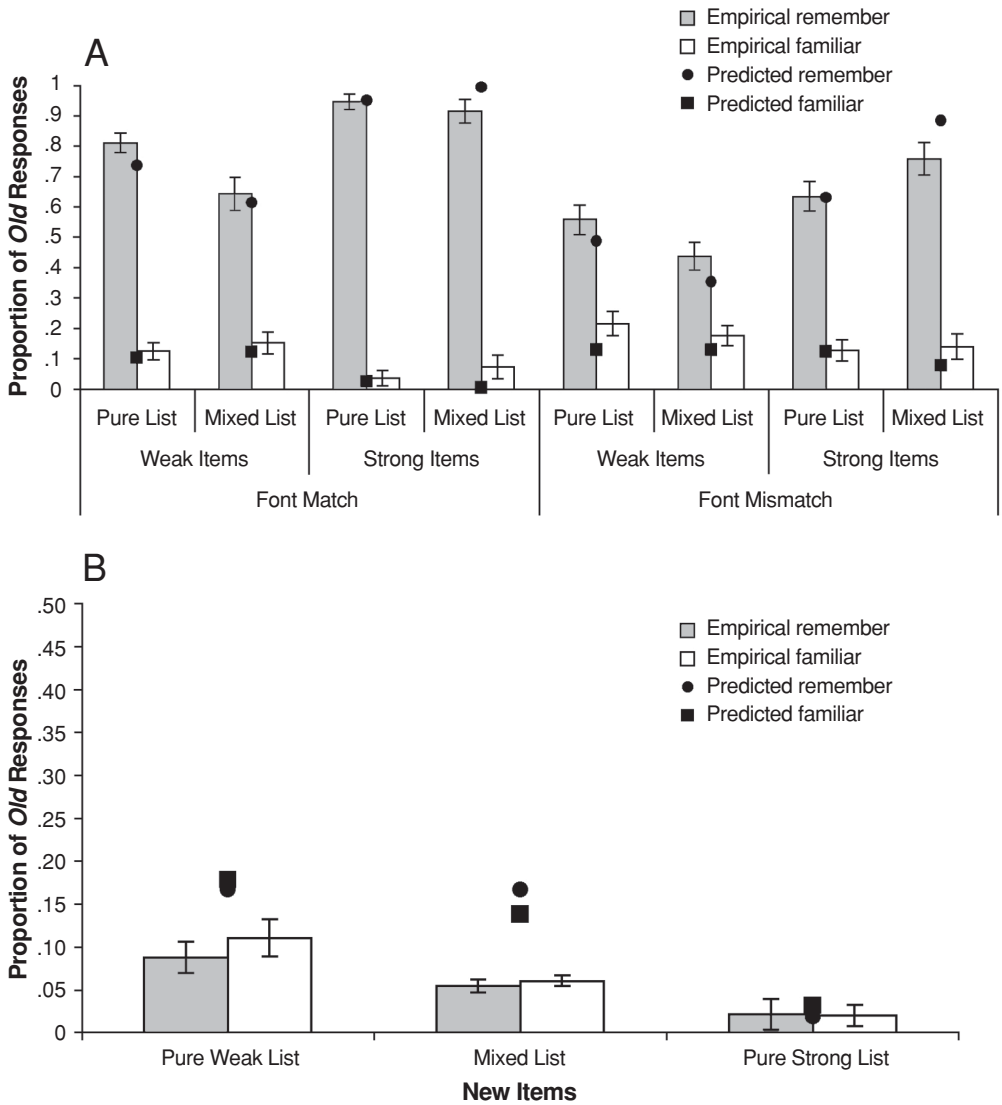


**Figure 5. Proportion of *remember* and *familiar* hits (A) and false alarms (B) as a function of match condition and list type in Experiment 2. Standard error bars are also plotted, along with the predicted data from the SAC simulation.**

those presented in a different font. There was no main effect of list type ($p = .08$). The only significant interaction was of list type × item strength [$F(1,32) = 17.16$, $p < .001$]. This interaction shows that weak items on the pure list were better recollected than those on the mixed list [$t(32) = 3.67$, $p < .01$], whereas recollection for strong items on the mixed list was not significantly different from that for those on the pure list [$t(32) = -1.53$, $p > .05$]. The lack of a significant difference between strong items on the mixed and the pure lists was at least partly due to the ceiling effects in the font match condition; therefore, we performed a *t* test on the font mismatch condition only for the strong items on the pure and mixed lists. This test revealed a trend toward the mixed list items being better recollected than the pure list items [$t(32) = -1.87$, $p = .07$].

As can be seen in Figure 5B, overall, there were more false alarms to the pure weak list than to the mixed list, with the pure strong list producing the fewest false alarms. We performed an ANOVA on the false alarms with factors of list type and response type (*remember* or *familiar*). We found a significant main effect of list type [$F(2,64) = 19.94$, $p < .001$] but no main effect of response type and no interaction.

We looked for an effect of $d'$ for *remember* and *old* responses, as can be seen in Figure 6. The ANOVA showed a significant effect of font match condition for both *remember* $d'$ [$F(1,32) = 67.59$, $p < .001$] and *old* $d'$ [$F(1,32) = 58.84$, $p < .001$]. Although we predicted that the differences in the font conditions would occur largely in the *remember* responses, it is not surprising to

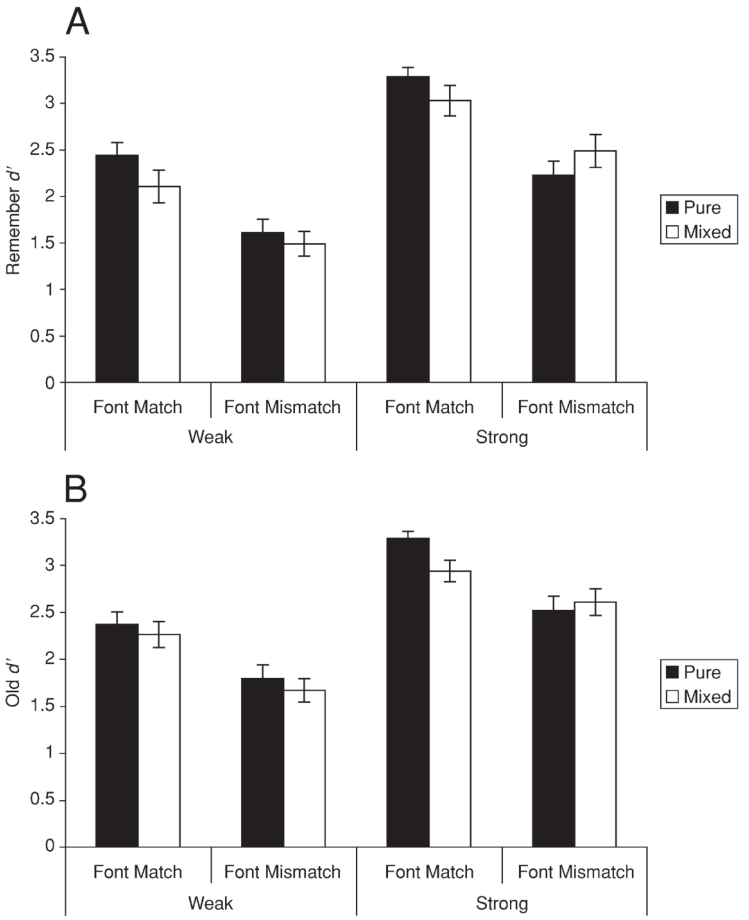


**Figure 6. Mean *d′* calculated with *remember* responses (A) and *old* responses (B) as a function of match condition and list type in Experiment 2.**

find effects when *familiar* responses are included as well. By including a font manipulation, we added an extra cue that would increase the participants' ability to recollect the episode from the study phase. Thus, the *old d′* has a larger influence from *remember* responses than it might in other experiments. Both ANOVAs also showed significant main effects of item strength [for *remember d′*, $F(1,32) = 84.27$, $p < .001$; for *old d′*, $F(1,32) = 67.00$, $p < .001$]. *Remember d′* also showed an interaction between list type and font match [$F(1,32) = 4.51$, $p < .05$], such that font match items produced a higher *d′* on the pure list than on the mixed list overall, whereas there was no difference between the pure and the mixed lists, overall, in the font mismatch condition.

The other main effects and interactions were not significant. This includes the interaction of list type and item type, which would indicate an LSE in *d′*. Visual inspection of the *remember d′* values shows that weak items were consistently better remembered for the pure list than for the mixed list, as we would predict. Strong items were better remembered for the mixed list than for the pure list, but only in the font mismatch condition. In the font match conditions, *d′* was at or above 3.0, which we believe is approaching the ceiling and, thus, may not accurately show differences between the pure and the mixed lists. Thus, although the *remember d′* effects were not statistically significant, they were in the correct direction. As in Experiment 1, the effects of list strength are seen in the *remember* hit rate, but not in the *remember d′*.

Experiment 1 showed a significant LSE for *remember* RTs. In order to determine whether this effect was replicated, we analyzed the false alarms in this experiment (see Table 2). Neither *remember* nor *familiar* responses showed a difference in RTs for the main effect of list type or the interaction of list type and item strength. However, RTs were shorter for words when the font at test matched the font used during study for both *remember* responses [$F(1,32) = 31.06$, $p < .001$] and *familiar* responses [$F(1,32) = 78.94$, $p < .001$]. RTs were also shorter to strong items than to weak items for *remember* responses [$F(1,32) = 15.64$, $p < .001$] and *familiar* responses [$F(1,32) = 59.28$, $p < .001$]. There was a significant interaction between item strength and font match for both *remember* RTs [$F(1,32) = 6.40$, $p < .05$] and *familiar*

**Table 2**
**Mean Reaction Times (in Milliseconds) for Experiment 2 as a Function of Response Type, List Type, Item Type, and Font Match**

| | Font Match | | Font Mismatch | |
|---|---|---|---|---|
| Response | Pure List | Mixed List | Pure List | Mixed List |
| Remember | | | | |
| Weak items | 1,258 | 1,312 | 1,464 | 1,304 |
| Strong items | 1,004 | 1,006 | 1,262 | 1,354 |
| Familiar | | | | |
| Weak items | 1,691 | 1,821 | 2,100 | 2,136 |
| Strong items | 1,110 | 1,188 | 1,954 | 1,838 |

RTs [$F(1,32) = 7.64$, $p < .01$]. This interaction shows that weak items were recognized more slowly than strong items when the font matched at study and test; however, there was no difference between the item strengths when the font was swapped. *Remember* RTs also showed a three-way interaction between list type, item strength, and font match [$F(1,32) = 5.55$, $p < .05$]. This interaction seems to show that for font match items, there was no difference in RTs for different list types; however, for font mismatch items, there was an interaction such that RTs for weak items on the pure list were longer than those for weak items on the mixed list, whereas strong items were not different in RT on the pure list versus the mixed list. Once again, the *familiar* responses were slower, overall, than the *remember* responses, as SAC would predict.

As in Experiment 1, we found that the effects of list strength were seen in the hit rate, whereas $d'$ effects were not significant, due to the increase in false alarms. Also, we were unable to replicate our RT findings in Experiment 1. Once again, we feel that an analysis of the *familiar* hit rate on each list will help to clarify whether any LSE may exist, in addition to the change in bias. The means of the *familiar* hit rates are very close, when compared across list type. Reflecting this, the three-way within-subjects ANOVA showed no significant effects of list type and no list type × item strength interaction. The only significant effects were the main effects of font match [$F(1,32) = 8.03$, $p < .01$] and item strength [$F(1,32) = 9.66$, $p < .01$]. The fact that the *remember* hit rates on pure and mixed lists differed (in the three conditions in which they were not at ceiling) by more than 12% in the correct direction, whereas the *familiar* hit rates never differed by more than 4% and in varying directions, indicates that bias may not be the only influence on *remember* responses.

## SAC MODEL FITS

Using the equations and specific processes of the SAC model, detailed in Reder et al. (2000), we modeled the data from Experiments 1 and 2. SAC predicts the percentage of recollection-based and familiarity-based responses that will be produced under the various conditions of a recognition task. These predicted response percentages are based on the current activation values of memory traces within the model. The relationships among those memory traces are represented in Figure 1.

It is assumed that all of the words used in Experiments 1 and 2 are already represented in memory in word nodes. Therefore, the nodes that need to be created during study are the context nodes and the episode nodes. The experimental context node represents those characteristics of the environment that the participant experiences during the experiment, such as the lighting, the equipment in the room, and the mood during the task. If the word is presented in a novel font, a font node is created to represent the perceptual qualities of the word. If the word is presented in a font that has been seen with another word, a link is made between the previously created font node and the new episode node. The word node, font node, and experimental context node are bound together by an episode node, which represents the experience of studying the word, with its perceptual characteristics, in the experiment.

When a probe word is presented at test, its corresponding node in memory is activated, along with the experimental context node and the font node. If the word is presented in the same font that was linked to the episode node during study, the font node will be a relevant source of activation that can spread to the episode node. The activation from the word and context nodes (font and experimental) may intersect at the same episode node, depending on whether the probe is a target item or a lure item and whether the same font is reinstated at test. Activation of episode nodes and word nodes produces recollection and familiarity-based judgments, respectively. That is, recollection responses are based on the activation of the episode node, where activation accrues due to spread from associated word nodes, font nodes, and experimental context nodes. Familiarity responses are based on the activation of the word node and, sometimes spuriously, on the activation of the font node.

The initial strength of each word node is based on the participant's history of exposure to that word, which is estimated on the basis of word frequency. This baseline activation ($B$) of a word both increases and decays slowly, according to a power function

$$B = B_{\mathrm{w}} + \ln\left(c_{\mathrm{N}} \sum t_i^{-d_{\mathrm{N}}}\right),$$

in which $B_{\mathrm{w}}$ is the base-level activation of the node (set to zero for episode nodes), $c_{\mathrm{N}}$ and $d_{\mathrm{N}}$ are constants, and $t_i$ is the time since the $i$th presentation. This is different from current activation $A$, which is higher than baseline whenever the word occurs in the environment or receives activation from its connections. The current activations for each word node, font node, and episode node were recalculated for every repetition of a strong item. Current activation decays according to an exponential function, moving back toward the baseline within a short period of time:

$$\Delta A = -\rho(A - B).$$

After each trial, current activation decreases by the proportion $\rho$ times the node's current distance from base-level activation. The variable $\rho$ is a stable parameter with a value of .8.

Activation spreads from each node in the structure that is activated by the environment (including the word node, font node, and experimental context node) to all other connected nodes. The activation spreads according to the number and relative strength of the links connected to the node, such that more links result in less activation spread along each individual link. For example, a word node may be connected to many episode nodes that bind the word to the various situations in which that word has been studied. The amount of activation that any node *r* receives is calculated according to the following equation:

$$\Delta A_r = \sum \left( A_s * S_{s,r} / \sum S_{s,I} \right),$$

in which $\Delta A_r$ is the change in activation of the receiving node, $A_s$ is the activation of each source node *s*, $S_{s,r}$ is the strength of the link between nodes *s* and *r*, and $\sum S_{s,I}$ is the sum of the strengths of all links emanating from node *s*. The total spread of activation is limited to the node *s*'s current activation.

The links between nodes vary in strength. This is based on the frequency and recency with which two pieces of information have been associated. The strengthening and weakening of these links occurs according to a power function

$$S_{s,r} = \ln\left( c_{\mathrm{L}} \sum t_i^{-d_{\mathrm{L}}} \right).$$

In this equation, $S_{s,r}$ is the strength of the link from node *s* to node *r*, $t_i$ is the time since the *i*th association between the two nodes, and $c_{\mathrm{L}}$ and $d_{\mathrm{L}}$ are constants for the links.

Once the activation of each node and its subsequent spread of activation to other nodes have been calculated, we can determine the probability of making a *remember* or a *familiar* response. The probability of a *remember* response is calculated by assuming a normal distribution of activation with fixed variance and a threshold for responding *remember*. The probability is computed by the following formula:

$$P(\mathrm{R}) = N\left[ \left( A_{\mathrm{E}} - T_{\mathrm{E}} \right) / \sigma_{\mathrm{E}} \right],$$

in which $A_{\mathrm{E}}$ is the activation of the episode node, $T_{\mathrm{E}}$ is the participant's threshold for the episode node, and $\sigma_{\mathrm{E}}$ is the standard deviation of the episode node's activation distribution. $N[x]$ is the area under the standard normal curve to the left of *x*. The probability of a *know* response is calculated as

$$P(\mathrm{F}) = \left\{ \left[ 1 - P(\mathrm{R}) \right] * N\left[ \left( A_{\mathrm{w}} - T_{\mathrm{w}} \right) / \sigma_{\mathrm{w}} \right] \right\} + \left\{ \left[ 1 - \left\langle P(\mathrm{R}) + P(\mathrm{F_w}) \right\rangle \right] * N\left[ \left( A_{\mathrm{f}} - T_{\mathrm{f}} \right) / \sigma_{\mathrm{f}} \right] \right\}.$$

In other words, the probability of a *familiar* response is the probability of not responding *remember* multiplied by the probability of the word node or font context node being above threshold. $P(\mathrm{F_w})$ represents the likelihood of a *familiar* response due to activation of the word node, when a *remember* response is not made. The f subscript indicates that the parameter is based on the font node. The effect of the activation from the font node on the probability of making a *familiar* response is important because font was varied between study and test, thus making it possible for participants to respond to lures on the basis of the familiarity of the font alone.

The fits to the list strength data are produced by the combination of two aspects of the model. The activation values of the episode nodes are greater for strong items and have a stronger link to the experimental context, leading to lower activations for weak items on a mixed list and higher activations for strong items on a mixed list (see Table 3). However, these differences in activation are small when compared with the total amount of activation of the episode nodes and word nodes. In order to fit the data well, different thresholds must be used for pure weak, mixed, and pure strong lists. This implies that the LSE is at least partially due to bias, although the model predicts that it is also partially due to competition among items.

Each model predicted *remember* and *familiar* responses for all the conditions in the experiments. The model in Experiment 1 had five free parameters representing the thresholds for the episode and word nodes on the mixed and pure lists, as well as the current boost. All parameter values are listed in Table 4. The model fit to the data was good (see Figure 3, comparing empirical and predicted data for hits and false alarms), with a sum of squared error value of .02 and an $R^2$ value of .95. We modeled Experiment 2 with six free parameters, representing thresholds for the episode and word nodes on each of the three lists. The resulting model fit was good (see Figure 5, comparing empirical and predicted data for hits and false alarms), with a sum of squared error value of .09 and an $R^2$ value of .96. Both of these models demonstrate that the predictions we claim are made by SAC do, in fact, fit the empirical data when the likelihood of making a *remember* or a *familiar* judgment is calculated on the basis of activation.

## GENERAL DISCUSSION

The present experiments were designed to test a dual-process account of recognition memory that makes specific predictions of hits and false alarms when list strength and context effects are manipulated. SAC's account of the LSE is based on activation spreading from reinstatement of the experimental context to the episodic memory trace. If all the words on the list are equally associated with the experimental context, they will receive equal activation from that context. However, if some words on the list have been studied more often, their links will be stronger than the words studied only once. This means that less activa-

**Table 3**
**Predicted Activation Values for the Event Nodes in the SAC Models in Experiments 1 and 2 as a Function of Item Type, List Type, and Font Match**

| | Font Match | | Font Mismatch | |
|---|---|---|---|---|
| | Pure List | Mixed List | Pure List | Mixed List |
| Experiment 1 | | | | |
| Weak items | 61.65 | 60.50 | 35.13 | 33.99 |
| Experiment 2 | | | | |
| Weak items | 94.10 | 92.35 | 57.51 | 55.76 |
| Strong items | 149.00 | 150.80 | 96.75 | 98.45 |

**Table 4**
**Parameter Values for SAC Models**

| Parameter | Value |
|---|---|
| $c_N$ | 25.0 |
| $d_N$ | 0.175 |
| $c_L$ | 25.0 |
| $d_L$ | 0.12 |
| $t_i$ | 50.3 |
| Current boost | 24.0 |
| Base-level word activation | 4.5 |
| $\sigma_e$ | 40.0 |
| $\sigma_w$ | 20.0 |
| Experiment 1, $T_e$ pure weak list | 40.4 |
| Experiment 1, $T_w$ pure weak list | 47.2 |
| Experiment 1, $T_e$ mixed list | 62.0 |
| Experiment 1, $T_w$ mixed list | 57.0 |
| Experiment 2, $T_e$ pure weak list | 38.7 |
| Experiment 2, $T_w$ pure weak list | 44.3 |
| Experiment 2, $T_e$ mixed list | 50.7 |
| Experiment 2, $T_w$ mixed list | 47.9 |
| Experiment 2, $T_e$ pure strong list | 83.5 |
| Experiment 2, $T_w$ pure strong list | 65.5 |

tion will spread to the episode nodes of the weaker words from the experimental context, making those weak words more difficult to recognize via the recollection process.

In order to test SAC's account of the LSE's being due to differential spread of activation affecting the recollection process, we included tests of the model's predictions regarding additional sources of activation—namely, the font of the word. The experiments described here demonstrate that reinstating the physical characteristics, or context, of a word at test will increase recognition accuracy. It should be noted that this effect of reinstating context is likely to occur only when the contextual information is integrated with the target information during encoding, as was encouraged in these experiments by the encoding task and the physical integration of the word and the font (Murnane, Phelps, & Malmberg, 1999).

Both experiments showed evidence of an LSE for recognition in *remember* responses, such that there are more *remember* responses to words when other words on the list provide less competition. In Experiment 1, this pattern was shown in hit rates and, for the first time, in RTs. Experiment 2 was designed to replicate and extend Experiment 1 and clearly showed an LSE in the hit rates as well. These LSE differences were fit by the SAC model, using a combination of activation differences and threshold differences.

Clearly the largest effect of list strength on recognition is a change in participants' bias to respond *old*. This change in bias has been discussed in previous articles on the LSE (Chappell & Humphreys, 1994; Hirshman, 1995; Murnane & Shiffrin, 1991a). We believe that changes in bias are manifested as threshold changes. That is, if participants are less confident about their memory but want to maintain a 50% rate of *old* responses, they will lower their thresholds such that a weaker activation of the episode will trigger a *remember* response and a lower amount of familiarity will trigger a *familiar* response. Therefore, when a pure weak list is tested, the overall memory for the list is worse than that for a mixed list, and participants will be forced to lower their thresholds to maintain responding at 50%. When a pure strong list is tested, memory will be better, overall, than on the mixed list, and the thresholds will be more conservative. This bias is particularly evident in our measurement of false alarms, where the pure weak list produces the most false alarms and the pure strong list produces the fewest false alarms. Measures of $d'$ are very sensitive to these changes in bias and, thus, do not show any differences after bias is removed. However, when our *familiar* responses were analyzed, they did not show the same changes in hit rate as the *remember* responses between the lists. We also showed an effect of list strength on the *remember* RTs in Experiment 1, which provides converging evidence that there is some interference at work between weak and strong items. However, because this effect did not replicate in Experiment 2, it may not be robust. We believe this provides at least tentative evidence that an effect is occurring in addition to bias. Our SAC model captures this combination of bias and interference.

It may be impossible to conduct a list strength experiment that does not show an effect of bias. It may also be true that this bias is hiding any small, but real, effects of list strength, as we argue. Norman's (1999, 2002) experiments show clear evidence of an LSE in recollection responses when the optimum conditions for recollection are presented. These findings provide evidence that there are real effects of list strength in recollection; however, it is becoming increasingly evident that the LSE is a very small effect. This is consistent with the SAC explanation that the only difference between weak items on a pure list and those on a mixed list is the competition for activation from the list context. This list context activation is only a small contributor to recollection. Perhaps a more effective way to test whether there is any influence of list strength is to create a mixed list on which only a few items are weak and the majority of the items are strong. In this situation, it is hard to imagine that the influence of the strong items would not make it more difficult to recognize the weak items than when all items are weak.

This is not the first study in which the effect of context on the LSE has been examined. As was described earlier, Murnane and Shiffrin (1991b) tested the effect of presenting words in different sentences at each repetition. They found that an LSE occurred for overall recognition in this case, but not when each repetition of the word was presented with the same sentence. The authors proposed that each repetition of the word was stored separately, due to the change in context at encoding, and thus the manipulation produced a change in list length. We agree with their interpretation of the result when words were repeated with different sentences. SAC would predict that increasing the number of sentences seen in a list is analogous to increasing the number of words on a list and, thus, creates an LLE, as was claimed by Murnane and Shiffrin (1991b). If modeled in SAC, the resulting change in recognition ability

would be due to an increase in the number of connections to the experimental context node, thus reducing the activation spreading to any single episode node. We disagree with their conclusion that there will be no evidence of an LSE when context remains the same for each repetition. We believe that the present experiments demonstrate evidence for a small LSE in recollection when words are re-presented in the same context.

Evidence continues to accumulate that suggests that the effect of adding items to a list (as seen in the traditional LLE and the study described above) is stronger than the effect of increasing the strength of some items on the list. The difference between these two effects in SAC is that the LLE involves an increase in the number of connections to the experimental context node, whereas the LSE involves an increase in the strength of some connections to the experimental context node. This implies that the number of connections to a node makes a larger difference than does the strength of those connections. The equations in SAC postulate that links, like nodes, increase in strength according to a power function. Therefore, each incremental increase in link strength should be less than the previous one, which is consistent with the notion that a greater number of links provide greater interference than does strengthening one link the same number of times.

**Comparison With Single Process Models of LSE**

Global matching models of recognition memory propose that items are recognized on the basis of a familiarity process. Originally, some of these models had predicted an LSE for recognition (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982). Although each model has different details, the reason for predicting an LSE is that the variance of the set of distractor items increases when some items are strengthened. Essentially, increasing the strength of items has the same outcome as increasing the number of items on the list. However, once the null LSE for recognition was clearly established, Shiffrin, Ratcliff, and Clark (1990) demonstrated that Gillund and Shiffrin's (1984) SAM model could be modified to predict a null LSE by adding a differentiation assumption. That is, as words on a list are strengthened, it becomes less likely that a lure will be confused with those words. Strengthening some items on a list does not impair memory for other items.

Other theorists also have included assumptions in their theories to explain the null LSE. Murdock and Kahana (1993) claimed that the variance in familiarity is only negligibly increased by stronger items, because the set of items studied in the experiment is so small, as compared with items studied over a lifetime. Dennis and Humphreys (2001) claimed that interference is more likely to occur due to exposure to the items outside of the experiment, rather than to increasing the strength of other items within the experiment. The Chappell and Humphreys model (1994) predicts a null LSE, because it claims that an episodic association is formed only at the first presentation of an item, regardless of study time.

There are at least two single-process theories that explain the null LSE without resorting to special assumptions or alterations. Both REM (Shiffrin & Steyvers, 1997) and the McClelland and Chappell (1998) model have differentiation instantiated in their processes. Within REM, stronger items have more features stored in memory and, thus, are less easily confused with a lure word. McClelland and Chappell have a similar approach, in that stronger items have a lesser degree of match with lure items and, therefore, contribute less likelihood. In both of these cases, strengthening of some items produces differentiation between those items and lure items, leading to a reduction in the number of false alarms but no detriment to weak items.

Models that predict a null LSE can account for a large portion of the list strength data. However, these models do not readily explain why *remember* responses, when analyzed separately, do show the LSE. Some researchers (Donaldson, 1996; Hirshman & Master, 1997) have proposed that *remember* and *know* responses are simply high-threshold and low-threshold responses, respectively, rather than reflecting fundamentally different processes. If this were true, it is not obvious why single-process models might predict that high-threshold responses would show an LSE, whereas overall responding does not. We claim that *remember* and *know* responses do, in fact, represent the operation of two separate processes (see Rotello, Macmillan, & Reeder, 2004) and explain the LSE according to the characteristics of these two processes. The single process described in this section does not provide for a recollection process.

It is possible that the REM and the McClelland and Chappell (1998) model could accurately model results that show a recollection LSE, as in Norman's (1999, 2002) experiments. However, the articles establishing these models (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997) claim that they predict a null LSE. Thus, even if these models were shown to numerically match positive LSE results, it would seem that the fundamental predictions of the model do not agree with this finding. These models would probably have little difficulty fitting the data from our experiments; however, they might have difficulty explaining why our *remember* hit rate (which they might think of as high confidence) would show differences between lists, whereas our *familiar* hit rate (which they might think of as low confidence) does not. It seems intuitive that low-confidence responses would be *more* likely to show an influence of bias than would high-confidence responses.

Those single-process models that predict an LSE for cued recall could account for Norman's recollection LSE if they were amended such that *remember* responses are produced by a process more similar to cued recall than to a familiarity process. For example, Shiffrin et al. (1990) stated that their improved version of SAM will predict a moderate LSE for cued recall if it is assumed that the weight given to the context cue is higher in cued recall than in recognition. Also, incorrectly retrieving an item will strengthen the association between that item and the

context, and this will occur more often for strong items that are sampled more frequently. Therefore, the strong items are more strongly associated with the context, and the context is more important in cued recall than in recognition. If cued recall is mapped onto recollection and recognition, in general, is mapped onto familiarity, this proposal is somewhat similar to SAC.[3] In SAC, recollection requires activation of episodic information, which often involves spreading activation from the representation of the environmental context, whereas familiarity requires only direct activation of conceptual information. However, by proposing that the cued recall process in SAM be used to explain recollection, we have fundamentally changed the assumptions of the model, such that it becomes a dual-process model.

Norman and O'Reilly (2003) have also proposed a dual-process model that predicts the recollection LSE. Their model is implemented within the connectionist complementary learning systems framework. In particular, their model is helpful for understanding how a dual-process model of recognition could be implemented in the brain. The Norman and O'Reilly model also provides a framework within which to understand the recollection LSE, but we do not wish to speculate about what the predictions would be for their model concerning the context and fan manipulations used in our experiments.

## Summary

The recent finding that list strength impairs recollection but not familiarity provides support for a dual-process account of recognition. SAC provides a simple model that explains the mechanisms behind the LSE, as well as context effects. Both effects occurred for *remember* hit rates and are well fit by predictions generated from an implemented version of this theory. No $d'$ effects were found, which we believe is due to the differences in bias between pure weak lists, mixed lists, and pure strong lists. A review of single-process models suggests that there is, as yet, no clear reason that they would predict the positive LSE in recollection, as demonstrated by Norman (1999, 2002).

## REFERENCES


Anderson, J. R. (1981). Interference: The relationship between response latency and response accuracy. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 326-343.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory & Language*, **49**, 231-248.

Chandler, C. C. (1991). How memory for an event is influenced by related events: Interference in modified recognition tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 115-125.

Chappell, M., & Humphreys, M. S. (1994). An auto-associative neural network for sparse representations: Analysis and application to models of recognition and cued recall. *Psychological Review*, **101**, 103-128.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments, & Computers*, **25**, 257-271.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, **108**, 452-478.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, **24**, 523-533.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1355-1369.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.

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

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. *Memory & Cognition*, **25**, 345-351.

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

Kim, K., & Glanzer, M. (1995). Intralist interference in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1096-1107.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, **105**, 724-760.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **90**, 316-338.

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

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, **128**, 403-415.

Murnane, K., & Shiffrin, R. M. (1991a). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 855-874.

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

Norman, K. A. (1999). Differential effects of list strength on recollection and familiarity (Doctoral dissertation, Harvard University). *Dissertation Abstracts International: Section B. The Sciences & Engineering*, **60**, 2974.

Norman, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1083-1094.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, **110**, 611-646.

Ohrt, D. D., & Gronlund, S. D. (1999). List-length effect and continuous memory: Confounds and solutions. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson–Shiffrin model* (pp. 105-125). Mahwah, NJ: Erlbaum.

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

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). 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, & Cognition*, **26**, 294-320.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum–difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review*, **111**, 588-616.

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

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology*, **92**, 297-304.

Willis, G. B., & Underwood, B. J. (1983). A lack of interference effects in recognition memory. *Bulletin of the Psychonomic Society*, **21**, 427-430.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1341-1354.

Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 345-355.

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, **48**, 516-534.

## NOTES

1. The amount of activation that spreads from the context node to a weak item from a pure weak list is effectively the same as the amount that spreads to a strong item from a pure strong list. The amount of activation that spreads is relative to the competition; however, strong items are better recollected than weak items, because the episode node is stronger due to repeated presentations, as is the link from the word node to the episode node.

2. SAC proposes that participants first attempt to recollect before assessing familiarity, and thus familiarity, although a faster process when performed alone, produces slower responses.

3. Thanks to Ken Norman for pointing this out.

(Manuscript received September 15, 2003;
revision accepted for publication October 24, 2004.)