

Strong memories obscure weak memories in associative recognition

MICHAEL F. VERDE and CAREN M. ROTELLO
University of Massachusetts, Amherst, Massachusetts

The list strength effect, in which strengthening some memories has a detrimental effect on the retrieval of other memories, has generally not been found in item recognition. The present study shows that the list strength effect does occur in associative recognition. Study materials were sets of overlapping word pairs (A-B, A-C, D-B, etc.). Within critical sets of words, strong pairs were presented three times at study, as compared with one presentation for weak pairs. In Experiment 1, associative recognition for weak pairs was less accurate than that for baseline pairs, and response times for hits were slower. In Experiment 2, receiver-operating characteristic curve data provided further evidence of poor accuracy for weak pairs. These findings support a qualitative distinction between item and associative recognition.

Do strong memories interfere with the retrieval of weaker memories? There has been little evidence for such interference in recognition memory (Murnane & Shiffrin, 1991a, 1991b; Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Huber, & Marinelli, 1995; Yonelinas, Hockley, & Murdock, 1992), with two notable exceptions. Murnane and Shiffrin (1991b) observed a list strength effect when individual words were repeated in the context of different sentences, but not when they were repeated within identical sentences. They argued that different-context repetitions result in the formation of separate memory representations. This effectively increases the length of the study list in the different-context condition, and it is well-known that list length interferes with recognition. In contrast, they suggested that identical-context repetitions lead to the strengthening of a *single* representation and that, under these conditions, there is no list strength effect.

Norman (2002), on the other hand, did observe a list strength effect in item recognition, using identical repetitions as a means of strengthening. He failed to find a reliable effect in a second experiment but did observe a list strength effect when subjects judged whether the plurality of recognition probes had changed from study to test. Plurality discrimination has been shown to involve recollective processing (Hintzman, Curran, & Oppy, 1992; Rotello, Macmillan, & Van Tassel, 2000; Westerman, 2001). Accordingly, Norman framed his results in terms of dual-process theory: Strength-based interference affects the *recollection* of specific episodic or associative memory but does not affect nonspecific *familiarity*. The

list strength effect may be elusive in recognition because the contribution of recollection is not typically isolated or controlled. The fact that only the plurality judgments produced a list strength effect in the second experiment is consistent with this conclusion.

In the present study, we were interested in associative recognition, which requires discriminating intact from rearranged word pairs. That memory for specific associations is required by the task suggests the predominant role of recollection, and this intuition is supported empirically by critical differences between associative and item recognition in time course (Doshier, 1984; Nobel & Shiffrin, 2001; Rotello & Heit, 1999, 2000), receiver-operating characteristic (ROC) curves (Rotello et al., 2000; Yonelinas, 1994, 1997), and susceptibility to some manipulations (for a review, see Clark & Gronlund, 1996). If strength-based interference is specific to recollection, as is suggested by its presence in recall (Tulving & Hastie, 1972; Wixted, Ghadisha, & Vera, 1997) and plurality recognition (Norman, 2002), a list strength effect should be readily observed in associative recognition. Such a finding would lend further support to the view that item and associative recognition rely on qualitatively different retrieval processes.

EXPERIMENT 1

The study list consisted of word pairs belonging to a number of overlapping sets. For example, *lake-monkey*, *lake-chest*, and *bread-monkey* might be members of the same set. In the baseline condition, all eight pairs in a set were studied once during study. In the critical condition, half of the pairs in a set were studied once (weak condition), and half were studied three times (strong condition). The crucial comparison is between baseline pairs and weak pairs. If the strength of related pairs has no effect on recognition, performance should not differ between

This research was supported in part by NIH Research Grant MH60274-02 to C.M.R. M.F.V. was supported by NIH Training Grant MH16745-19. Correspondence concerning this article should be addressed to M. F. Verde, Department of Psychology, Box 37710, University of Massachusetts, Amherst, MA 01003-7710 (e-mail: mverde@psych.umass.edu).

these two conditions. On the other hand, if strengthening some pairs in a set makes memory for weaker pairs less accessible, one would expect decreased accuracy and, perhaps, increased response time (RT) for the weak pairs, as compared with the baseline pairs. To decrease the likelihood that the subjects might differentially rehearse strong and weak items, they were not warned of the memory test. Instead, the word pairs were studied in the guise of an imagery-rating task.

Method

Subjects. Forty undergraduates from the University of Massachusetts at Amherst participated in the study for course credit.

Materials and Design. The subjects were assigned to individual computers, which controlled list generation, stimulus presentation, and response recording. The assignment of items to experimental conditions and the order of items within lists were uniquely randomized for each subject. The stimuli were drawn from a pool of 156 six-letter, low-frequency nouns.

Each *interference set* of word pairs was created from six unique words. Assigning three words to the left position and three to the right position allowed nine possible word pairings (A–D, A–E, A–F, B–D, B–E, B–F, C–D, C–E, and C–F), of which eight were chosen to make up the interference set. Nine sets were created (72 unique word pairs in all). Six sets were assigned to the critical condition, and three sets to the baseline condition. In the critical condition, half of the pairs in a set (A–E, B–D, C–D, and C–F) were presented three times at study (strong condition), whereas the other half (A–D, A–F, B–F, and C–E) were presented once (weak condition). Each individual word appeared in both strong and weak pairs within the set, so strength was defined at the level of the word pair or association. In the baseline condition, all the pairs in a set were presented once during study.

The study list consisted of 120 randomly ordered pairs. The test list consisted of 144 randomly ordered pairs: 72 unique pairs from the study list, as well as 72 rearranged pairs. Rearranged pairs were created by pairing two words from different sets but belonging to the same condition (strong, weak, or baseline).

Procedure. The 30-min session consisted of a study phase followed by a test phase. During the study phase, the subjects were not informed of the memory test to follow. They were simply asked to rate word pairs for imagery. On each trial, a single pair was presented on the screen for 2,000 msec, then was replaced by the prompt ??????. The subjects were instructed to form an image using both words and to respond as soon as the ?????? appeared. The subjects used the z and / keys to indicate whether it was easy or difficult to form a vivid image with the two words.

Following study, the subjects were given instructions for the surprise memory test, including a four-trial practice test. During each trial of the associative recognition test, a single pair was presented on the screen. The subjects used the z and / keys to indicate whether the probe was an intact or a rearranged pair.

Results

Test of the hypothesis that strengthening some pairs has a detrimental effect on memory for related pairs (those within the same set) required two comparisons for each dependent measure: strong condition versus baseline and weak condition versus baseline. An alpha of .05 was adopted for each comparison. Accuracy and RT data are shown in Table 1.

Recognition sensitivity (d') was calculated for each subject from hit and false alarm rates.¹ Sensitivity for the

Table 1
Experiment 1: Response Times (RTs) and Accuracy Data

Condition	d'	Hit Rate	FA Rate	Hit RT (msec)	CR RT (msec)
Strong	1.82	.81	.23	1,320	1,596
Weak	1.03	.58	.24	1,504	1,584
Baseline	1.27	.63	.22	1,413	1,624

Note—FA, false alarm; CR, correct rejection.

strong condition ($d' = 1.82$) was greater than that for the baseline condition ($d' = 1.27$; $t = 5.64$, $p < .001$). Sensitivity for the weak condition ($d' = 1.03$) was less than baseline ($t = 2.54$, $p < .05$).

The correct identification of old pairs was faster for the strong condition (1,320 msec) than for the baseline condition (1,413 msec; $t = 3.22$, $p < .01$). However, the hit RT was longer for the weak condition (1,504 msec), as compared with baseline ($t = 2.42$, $p < .05$). None of the conditions differed reliably with respect to the correct rejection of rearranged items. Miss and false alarm RTs were not analyzed, due to missing cells for some subjects.

Both the RT and the accuracy data show list strength effects in associative recognition. To our knowledge, this is the first report of such an effect.

EXPERIMENT 2

In Experiment 2, our goal was to replicate the findings of Experiment 1 while examining more closely the effect of list strength on accuracy. When there are different classes of test items, it is a concern that people may use different decision criteria for different classes. Under such conditions, accuracy measures based on binary old–new judgments (such as d') can be misleading (Donaldson, 1993; Verde & Rotello, 2003). In the present experiment, confidence ratings were collected and used to construct ROC curves, which provide a more detailed view of recognition performance and also allow the measurement of accuracy independently of response bias (Macmillan & Creelman, 1991). The procedure was otherwise identical to that in Experiment 1.

Method

Subjects. Thirty-one undergraduates from the University of Massachusetts, Amherst, participated in the study for course credit.

Materials and Design. The materials and design were identical to those in Experiment 1.

Procedure. The procedure was identical to that in Experiment 1, except that the binary (intact–rearranged) recognition judgment was replaced by a confidence rating on a 6-point scale (6 = *very confident intact*, 1 = *very confident rearranged*).

Results

Recognition performance is shown in Table 2. Hit and false alarm rates were derived by combining confidence ratings 1–3 to form the *rearranged* response category and ratings 4–6 to form the *intact* response category. The confidence ratings were then used to construct an ROC curve for each experimental condition for each subject,

Table 2
Experiment 2: Accuracy Data

Condition	d_a	Slope	Hit Rate	FA Rate
Strong	1.89	0.80	.87	.24
Weak	1.05	0.79	.65	.24
Baseline	1.28	0.71	.70	.20

Note—FA, false alarm.

using maximum likelihood estimation. The slope of the z -transformed ROC curve (z slope), combined with normalized hit and false alarm rates, provides a measure of sensitivity (d_a) that takes account of the variances of the distributions (measured with z slope) and is independent of response bias (Macmillan & Creelman, 1991):

$$d_a = [2 / (1 + z\text{slope}^2)]^{1/2} [z(H) - z\text{slope} \cdot z(\text{FA})], \quad (1)$$

where z is the inverse normal function.

The use of this alternative measure of sensitivity is important in this experiment, since the z -transformed ROC slopes were not equal to 1. Recognition sensitivity for the strong condition ($d_a = 1.89$) was greater than that for the baseline condition [$d_a = 1.28$; $t(31) = 7.05$, $p < .001$]. Sensitivity for the weak condition ($d_a = 1.05$) was less than baseline [$t(31) = 3.01$, $p < .01$]. Thus, even when we use a measure that adjusts for different z ROC slopes and decision criteria, the list strength effect is observed. These differences in sensitivity are graphically evident in the ROC curves for each condition (Figure 1): The strong condition ROC falls above the baseline condition ROC, which falls above the weak condition ROC.

We also noted that the slope was marginally greater for the weak condition (slope = 0.79) than for the baseline condition [slope = 0.71; $t(31) = 3.01$, $p < .10$], al-

though the difference between the strong (slope = 0.80) and the baseline conditions was not reliable. The theoretical significance of this slope difference is unclear, and indeed, explaining empirically observed changes in slope remains a weakness for most recognition models (Heathcote, 2003).

GENERAL DISCUSSION

In the present study, we found a robust list strength effect in associative recognition. Strengthening some pairs via repetition had an adverse effect on both RT and accuracy for related nonrepeated pairs. Studies in which item recognition has been looked at have almost uniformly failed to show such an effect. Only Norman (2002) observed a list strength effect in item recognition, and he found the effect difficult to replicate.

Ratcliff et al. (1990) noted that the null findings were a problem for a number of familiarity-based recognition models, which describe familiarity as based on a global match between a probe and the contents of memory. If making some items in memory stronger leads to a greater match not only to identical probes, but also to mismatching probes, the resulting increase in noise (the variability of familiarity distributions) drives down sensitivity. Thus, the prediction of a list strength effect: For weak items studied with strong items, increased noise at retrieval leads to worse accuracy than that for weak items studied only with other weak items.

Subsequent models have been developed to account for the null list strength effect. However, the present results present a challenge for single-process models that assume that item and associative recognition rely on the same mechanism. For example, Ratcliff et al. (1990) pro-

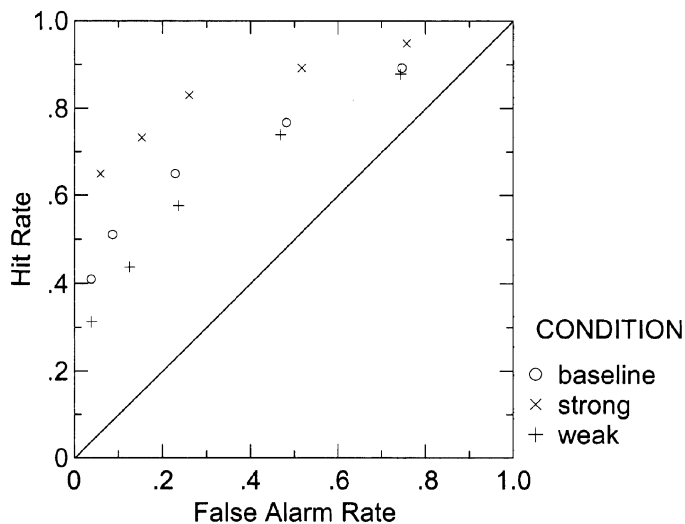


Figure 1. Receiver-operating characteristic data from the three conditions of Experiment 2. Note that memory performance is higher when the hit rate is higher for a given false alarm rate, so curves that fall higher in the space indicate better discrimination.

posed a modification of the SAM model in which repetition leads to differentiation, so that repeated items do not cause an increase in global match for mismatching probes (an idea that also appears in other models). Although the differentiation parameters might be flexible enough to accommodate a list strength effect in some instances and not others, there should be a compelling argument for why item and associative recognition require different parameter values.

An alternative to the single-process approach is to assume that multiple component processes underlie recognition, with the contribution of these components varying across tasks. Intuitively, recollection of specific information, rather than global familiarity, should predominate in associative recognition for functional reasons: The familiarity of individual words does not help one to discriminate intact from rearranged pairs. This is supported empirically by findings that associative recognition is not very susceptible to manipulations thought to target familiarity (Cameron & Hockley, 2000; Westerman, 2000, 2001) but is particularly diminished in older adults, whose memory deficits may be primarily recollective in nature (Naveh-Benjamin, 2000). This may be why associative recognition often resembles recall rather than item recognition, as in the present findings (for other examples, see Clark, 1992; Clark & Burchett, 1994; Nobel & Shiffrin, 2001; Verde, in press).

The use of recall in a recognition task is often invoked to explain aspects of data not sufficiently accommodated by a familiarity-based model (e.g., Clark, 1999; Clark, Hori, & Callan, 1993; Mandler, 1980; Shiffrin et al., 1995). Might recall serve as a model for recollection in recognition? Recall is typically described as a probabilistic sampling process, and in some models, interference can occur because the presence of strong memory items makes the sampling of weaker items less likely (Rohrer, 1996; Wixted et al., 1997; see also Raaijmakers & Shiffrin, 1981; Rundus, 1973). On the surface, such models predict patterns of interference in both accuracy and RT similar to those observed here. However, given the difference in quality of the retrieval cues available in recall and recognition, as well as the issue of how familiarity might interact with recollection, it is unlikely that models of recall can be adapted without considerable modification. A useful step would be to directly compare cued recall and associative recognition. Ratcliff et al. (1990) found the list strength effect in cued recall to be small and somewhat unreliable. This was not the case for associative recognition in the present study, but our use of overlapping sets of pairs made for a very strong manipulation of interference.²

OTHER ISSUES

The present findings are also relevant to recent debates concerning the theoretical significance of the shapes of ROC curves. Yonelinas's (1994) dual-process model of

recognition describes recollection as a single all-or-none threshold, which implies that tasks relying primarily on recollection should produce ROC curves that are more or less linear. Although such curves have been observed (Rotello et al., 2000; Yonelinas, 1994, 1997), it is clear from the present findings and others that both source and associative recognition commonly produce ROC curves that are unquestionably curvilinear (Kelley & Wixted, 2001; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick, Klein, Dodson, & Shimamura, 2000). This is strong evidence against recollection as a single-threshold process (cf. Malmberg, 2002), but there is no reason that recollection has to be characterized in this way. Kelley and Wixted (2001), for example, have described a *some-or-none* model in which both item and associative information are continuously distributed variables that can produce both relatively linear and curvilinear ROCs.

A final point of interest is that in both experiments, the effect of list strength was observed only in the hit rates, with the false alarm rates of the strong, weak, and baseline conditions being statistically indistinguishable. Kelley and Wixted (2001) also found in associative recognition that repeating word pairs increased hits but had no effect on false alarms. These results are puzzling given other evidence that the familiarity of individual words can influence associative recognition (Cleary, Curran, & Greene, 2001; Kelley & Wixted, 2001). If this were true, one would expect an increase in false alarms with repetition. Kelley and Wixted suggested that repetition not only made individual words more familiar, but also increased the likelihood of retrieving the original studied pair so that it could be used to reject the rearranged lure. These opposing tendencies canceled each other, leaving no apparent change in the false alarm rate. This explanation seems reasonable considering the abundant evidence for recall-to-reject strategies in recognition (e.g., Cleary et al., 2001; Hintzman et al., 1992; Jacoby, Jones, & Dolan, 1998; Rotello & Heit, 2000; Rotello et al., 2000). However, it is quite a coincidence that the effect of increasing familiarity would exactly cancel the effect of increasingly likely recall-to-reject across experiments that differed in design and materials. Future work in which varying the relative magnitude of these opposing tendencies can be shown to have different effects on the false alarm rate would make this explanation more persuasive.

REFERENCES

- CAMERON, T. E., & HOCKLEY, W. E. (2000). The revelation effect for item and associative recognition: Familiarity versus recollection. *Memory & Cognition*, *28*, 176-183.
- CLARK, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, *20*, 231-243.
- CLARK, S. E. (1999). Recalling to recognize and recognizing to recall. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model* (pp. 215-243). Mahwah, NJ: Erlbaum.
- CLARK, S. E., & BURCHETT, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition*, *22*, 55-62.

- CLARK, S. E., & GRONLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3**, 37-60.
- CLARK, S. E., HORI, A., & CALLAN, D. E. (1993). Forced-choice associative recognition: Implications for global-memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 871-881.
- CLEARY, A. M., CURRAN, T., & GREENE, R. L. (2001). Memory for detail in item versus associative recognition. *Memory & Cognition*, **29**, 413-423.
- DONALDSON, W. (1993). Accuracy of d' and A' as estimates of sensitivity. *Bulletin of the Psychonomic Society*, **31**, 271-274.
- DOSHER, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, **16**, 519-555.
- HEATHCOTE, A. (2003). Item recognition and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1210-1230.
- HINTZMAN, D., CURRAN, T., & OPPY, B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 667-680.
- JACOBY, L. L., JONES, T. C., & DOLAN, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review*, **5**, 705-709.
- KELLEY, R., & WIXTED, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 701-722.
- MACMILLAN, N., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MALMBERG, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 380-387.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- 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.
- NAVEH-BENJAMIN, M. (2000). Adult age difference in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1341-1354.
- NOBEL, P. A., & SHIFFRIN, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 384-413.
- NORMAN, K. A. (2002). Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1083-1094.
- QIN, J., RAYE, C. L., JOHNSON, M. K., & MITCHELL, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1110-1115.
- RAAJMAKERS, J. G. W., & SHIFFRIN, R. M. (1981). Search of associative memory. *Psychological Review*, **88**, 93-134.
- 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.
- ROHRER, D. (1996). On the relative and absolute strength of a memory trace. *Memory & Cognition*, **24**, 188-201.
- ROTELLO, C. M., & HEIT, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory & Language*, **40**, 432-453.
- ROTELLO, C. M., & HEIT, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, **28**, 907-922.
- ROTELLO, C. M., MACMILLAN, N. A., & VAN TASSEL, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language*, **43**, 67-88.
- RUNDUS, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior*, **12**, 43-50.
- SHIFFRIN, R. M., HUBER, D. E., & MARINELLI, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 267-287.
- SLOTNICK, S. D., KLEIN, S. A., DODSON, C. S., & SHIMAMURA, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1499-1517.
- TULVING, E., & HASTIE, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology*, **92**, 297-304.
- VERDE, M. F. (in press). Associative interference in recognition memory: A dual-process account. *Memory & Cognition*.
- VERDE, M. F., & ROTELLO, C. M. (2003). Does familiarity change in the revelation effect? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 739-746.
- WESTERMAN, D. L. (2000). Recollection-based recognition eliminates the revelation effect in memory. *Memory & Cognition*, **28**, 167-175.
- WESTERMAN, D. L. (2001). The role of familiarity in item recognition, associative recognition, and plurality recognition on self-paced and speeded tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 723-732.
- WIXTED, J. T., GHADISHA, H., & VERA, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 523-538.
- 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. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, **26**, 747-763.
- 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.

NOTES

1. Hit and false alarm rates were capped at .98 and .02, respectively, when d' was calculated. Capping hit and false alarm values at $(N + .5)/(N + 1)$ and $.5/(N + 1)$, respectively, is commonly done to avoid floor and ceiling values that can distort the range of d' (Macmillan & Creelman, 1991). About 5% of the data points were affected by this correction.

2. Norman (2002) similarly noted that the strength of his manipulation might be partly responsible for his finding a list strength effect in item recognition.

(Manuscript received August 19, 2003;
revision accepted for publication January 14, 2004.)