Journal of Experimental Psychology:
Learning, Memory, and Cognition
1993, Vol. 19, No. 6, 1445–1449

0278-7393/93/$3.00

TODAM and the List-Strength and List-Length Effects: Comment on Murdock and Kahana (1993a)

Richard Shiffrin, Roger Ratcliff, Kevin Murnane, and Peter Nobel

B. B. Murdock and M. J. Kahana (1993a) presented a continuous memory version of the theory of distributed associative memory (TODAM) model; they claimed that this model predicts list-strength and list-length findings, including those reported by R. Ratcliff, S. E. Clark, and R. M. Shiffrin (1990) and K. Murnane and R. M. Shiffrin (1991a). This model is quite similar to one discussed by R. M. Shiffrin, R. Ratcliff, and S. Clark (1990), who rejected the model on the basis of its inability to predict both an absent or negative list-strength effect (when strength is varied by repetitions) and a present list-length effect. In this comment we elaborate the earlier discussion and demonstrate that the version of TODAM proposed by B. B. Murdock and M. J. Kahana (1993a) indeed fails for this reason. We show this first for a somewhat simplified version of the model for which derivations are obvious and then in a simulation of the complete version using the parameter values suggested by B. B. Murdock and M. J. Kahana (1993a).

Murdock and Kahana (1993a) proposed a continuous memory variant of the theory of distributed associative memory (TODAM). Shiffrin, Ratcliff, and Clark (1990) discussed and rejected an almost identical version of this model. In this comment we elaborate our earlier discussion and show that this variant of TODAM fails to predict key elements of our data and of other data from the literature. To set the stage, we quote from Shiffrin et al., (1990, where they discuss the case of spaced repetitions): "A trade-off resulting in no variance increase would apply both to list-length and list-strength effects, whereas the data exhibit opposing effects" (p. 187).

Ratcliff, Clark, and Shiffrin (1990), Shiffrin et al. (1990), and Murnane and Shiffrin (1991a) examined the list-length effect (LLE; termed *positive* when adding words to a list reduces performance for the remaining words) and list-strength effect (LSE; termed *positive* when strengthening some words on a list reduces performance for the remaining words). The LLE is found in recognition, cued recall, and free recall. However, the LSE is found only in free recall—It is small or missing in cued recall and is missing or even negative in recognition. The LLE can be reintroduced in recognition when the words being tested are studied in different sentence contexts (Murnane & Shiffrin, 1991b). These authors argued that such results were inconsistent with the TODAM model and not easy to reconcile with sensible variants of TODAM (e.g., Murdock, 1982).

Richard Shiffrin and Peter Nobel, Department of Psychology, Indiana University Bloomington; Roger Ratcliff, Department of Psychology, Northwestern University; Kevin Murnane, Department of Psychology, Pennsylvania State University.

This research was supported by National Institute of Health Grant 12717 to Richard Shiffrin.

Correspondence concerning this article should be addressed to Richard Shiffrin, Psychology Department, Indiana University, Bloomington, Indiana 47405. Electronic mail may be sent to shiffrin@ucs.indiana.edu.

We limit our discussion to the recognition findings. Recognition sensitivity is typically defined as

$$d' = \frac{[E(T) - E(D)]}{Var(D)^{1/2}}$$

where the expectations and variance refer to the variable used to make the recognition decision (in TODAM, the inner product of the test vector with the memory vector, i.e., the match of the test item to memory), and T and D refer to target test and distractor test. For the moment, consider an item tested at a fixed recency (fixed study–test lag); in TODAM, $E(T) - E(D)$ does not change for such an item when a list of items is increased in length or when the strength of nontarget items on a list is increased. Thus changes in predictions for d' reduce to the question: Does the variance change with list length or list strength? If the variance depends only on the items in the recent list, as suggested in earlier versions of TODAM, then it increases with both list length and list strength, thus producing both a positive LSE and a positive LLE, even for items at fixed recency, which is contrary to the data.

In their commentary, Murdock and Kahana (1993a) reiterated the suggestion put forth by Shiffrin et al. (1990) that the activation of memory involves items prior to the current list (an assumption designed to keep variance roughly constant regardless of the composition of the most recent list). In fact, every past item, including those "items" occurring prior to the experimental setting, is weighted by a factor α^j, where j is the lag from study to the start of the current test period. To obtain predictions for items from the current list at a fixed recency, for increasing list length or list strength, we must obtain predictions for the variance (since the numerator of d' is unchanged by changes of length or strength). The variance expressions can be worked out (and, were given in slightly different forms in Shiffrin et al., 1990, as well as in the present reply), but the mathematical complexities tend to obscure understanding. The essential nature of the model and the basis for the predictions are far easier to discern in a slightly simplified version of this model. Because the simplifications do not change the nature of the predictions or the conclusions in any essential way, we consider first this version of the continuous memory model.

In most of our studies, we varied strength by varying the number of spaced repetitions. In this case, all items in an experimental session are presented for a fixed study time. Thus we assume that these are each learned with an encoding probability, p. Memory prior to the experimental session is represented for convenience as a large number, H, of singly presented items, each learned with an encoding parameter, q. Our first simplification is the assumption that $p = q$. If we vary only list length in a session, then TODAM's memory consists of a virtually infinite sum of identical items, weighted by geometrically decreasing factors, α^j, where j is the item's recency. It is trivial to observe in this case that variance stays constant throughout the session, regardless of the length of the current list.

When repetitions occur, the situation is complicated by the fact that the repetitions are correlated with each other, thus producing an increase in overall variance in TODAM. This yields a positive LSE. Our second simplification is the assumption that the contributions to the variance for repetitions are as they would be for new items, as if repetitions were not correlated. In this case, variance remains constant throughout the session, regardless of the length or strength of the most recent list.

This is the essence of the continuous memory version of TODAM applied to the case of spaced repetitions. Since variance does not change with variations in list length or list strength, neither factor affects d', for singly presented items at a fixed recency. When one adds to the model the complications we have bypassed in this analysis, a positive LSE is predicted, as shown by Murdock and Kahana (1993a) but it is of such small magnitude that it would not be detected experimentally. The model would still mispredict those cases where a significantly negative LSE is observed, but this is a relatively minor objection. The crucial difficulty is the one raised by Shiffrin, et al. (1990): Neither an LLE nor an LSE is predicted; the data show these effects to differ qualitatively and quantitatively.

To produce an LLE, Murdock and Kahana (1993a) suggested averaging d' performance across the entire list, rather than focusing on items at a given recency. Because the model predicts geometrically decreasing performance with increasing study–test lag, the average across the list is lower for longer lists. There are two critical difficulties with this solution.

First, if one averages across positions for length variations, then one must also do so for repetition variations. Stronger lists have more presentations, and performance is predicted to fall off geometrically with presentations. Thus an average across the whole list causes TODAM to predict a positive LSE. Such a prediction is of course contrary to all the data we have collected in our various studies. In practice, matters are even worse: TODAM predicts LSEs to be at least as large as LLEs (because of the fact that, according to TODAM, repetitions are correlated). The simulation we report later in this reply illustrates this point.

Second, one can of course examine serial position functions and compare items at comparable lags. We have done this in a number of studies (e.g., Murnane & Shiffrin, 1991a, especially Experiment 4). The pattern of list-length and list-strength findings we have described holds when lag from study to beginning of test is held constant: LLEs appear, and LSEs do not appear. However, the studies in Murnane and Shiffrin (1991a) were not perfectly suited to test TODAM because the lags within the test sequence were not exactly controlled across conditions. Fortunately, an excellent test of the LLE with total lag controlled can be found in Murdock's own research (Experiment 3 of Ratcliff & Murdock, 1976). List lengths of 16, 32, and 64 items were used. (For the 64-item lists, only the last 16 studied items and the first 32 items tested are used in the analyses below.) Figure 1 gives accuracy (d') as a function of the lag between study and test for an item, for the three list lengths. According to the modified TODAM model, the three lag functions should lie on top of one another. However, the lag functions disconfirm these predictions: The longer lists produce lower performance than the shorter lists for items at equal study–test lags.

To this point in our commentary we have used the simplified model to clarify the predictions of the continuous memory version of TODAM and to highlight the critical misprediction concerning the relation of the LLEs and LSEs. These results confirm the conclusions reached by Shiffrin et al. (1990). Is there anything in the complete and more complex model that would throw these conclusions into doubt? We briefly discuss the complications and then demonstrate the validity of our conclusions by simulating the full version of the model, using the parameter values suggested by Murdock and Kahana (1993a).

The full model introduces several complicating factors. First, q is assumed to be larger than p. By itself, this would produce a larger variance at the start of the session, a variance that gradually drops to a level lower by a factor of p/q. Second, the preexperimental items are assumed to have been singly presented, thus producing a lower variance in comparison with the session items because many session items are correlated repetitions; this increases the variance. This factor acts to increase the variance as the session continues, thus offsetting the first factor. Third, the structure of successive lists matters. This is especially true when counterbalancing is used, because a current list that tends to increase variance (e.g., many repetitions) is preceded by another list that tends to decrease variance (e.g., fewer repetitions) and vice versa. Thus, counterbalancing generally acts to equalize variance as additional lists are presented.

To deal with the LSE, Murdock and Kahana (1993a) assumed that q (.7) is greater than p (.5 or .3), that there is a high value of α (.995), that there is counterbalancing of successive list types, and that the list lengths are 20. Under these conditions, the three factors balance, variance stays fairly constant, and the predicted LSE is extremely small. However, the predicted LLE is not only extremely small but in the wrong direction (as shown later).

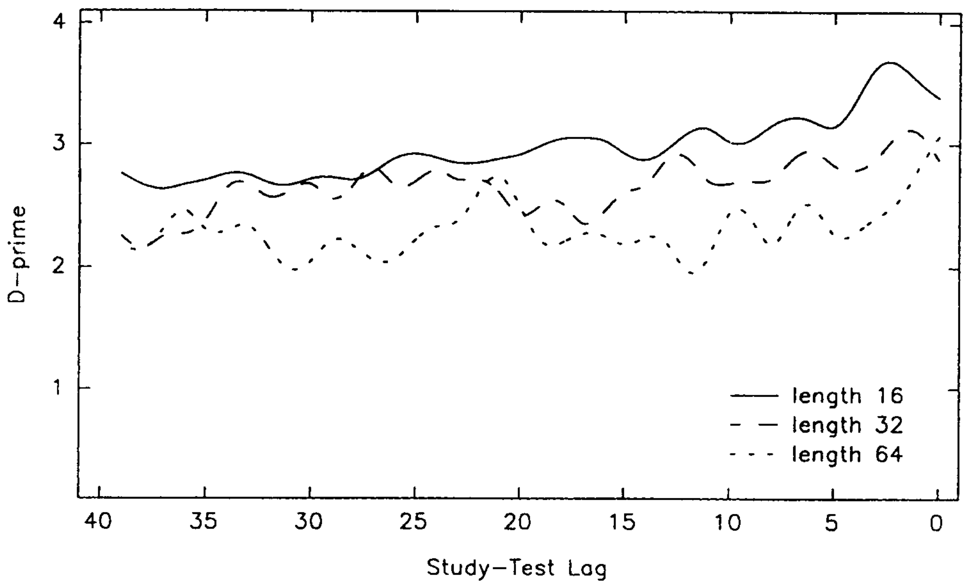

Figure 1. Data are from Ratcliff and Murdock (1976): recognition performance (d') as a function of study–test lag measured in intervening items, for lists of different lengths. (The false-alarm rate for each list length was the average for the first 32 test positions. The data have been smoothed slightly for convenience and clarity [if $x(i)$ is the actual data point at lag i, we graph $y(i) = .25x(i - 1) + .5x(i) + .25x(i + 1)$, except that $y(1) = .75x(1) + .25x(2)$ and $y(40) = .75x(40) + .25x(39)$].)

To verify the predictions of the full, new version of TODAM, we simulated the complete model with the parameters suggested by Murdock and Kahana (1993a), for a paradigm like that of Murnane and Shiffrin (1991a): There are eight lists, two each of four conditions: *Pure–Strong* (three repetitions of 10 items), *Pure–Weak* (one repetition of 10 items), *mixed* (5 singly presented items, 5 thrice-presented items), and *long* (40 singly presented items). The eight lists are used once each, in a random order for each simulated session. Ten targets and 10 distractors are tested for each list.

The simulation conformed to that proposed by Murdock and Kahana (1993a): The vectors each consisted of 1,000 features. We assumed an infinite series of preexperimental items presented once each, with $q = .7$. (Actually we started each simulation by filling a vector with random entries corresponding to the theoretical power calculated by Murdock and Kahana, and we confirmed that expression by a separate pilot simulation.) The predictions we give are based on 5,000 simulated sessions.

The simulation produces distributions of match values for each of the four conditions and for each of the eight positions in the session that a list–condition can occupy. For once-presented items, serial input position is well defined, and we have obtained simulated distributions at each list position. For items presented several times, serial position is less clearly summarized, so we give our serial position results for singly presented items only.

There are a number of conceivable measures of performance. We give predictions in terms of d', which is defined as the difference of the means of the predicted match distributions for targets and distractors divided by the standard deviation of the distractor match distribution (our findings do not differ in character for several other measures of performance that were examined).

The number of simulated sessions we used should guarantee that the simulation results are very close to the theoretical predictions of the model. (In several different runs, the results we report did not vary appreciably.)

The most important results from the simulation are given in Figure 2. Performance (d') averaged across the session is plotted for weak items as a function of serial position. There are very small differences among the conditions. In fact, at corresponding serial positions, the longer list actually is superior to the shorter lists, unlike our observed data (and all other data), which exhibit poorer performance for longer lists. In addition, the strength manipulation had very little effect. Averages taken across only the 10 most recent serial positions are 1.31, 1.31, and 1.36 for the pure weak, mixed weak, and long lists respectively, with the difference between 1.36 and the others being statistically significant by a test across the 10 serial positions, $t(9) = 7.24$, $p < .001$.

Averaging all serial positions together mixes in the lower performance items in the earlier positions of the longer and stronger lists and lowers performance for these lists. As a result average performance shows small positive list-length and list-strength effects: Averaged across the entire list, d' is 1.31 for pure weak lists, 1.28 for mixed weak lists, and 1.27 for long lists. This analysis is important only because it shows that even the incorporation of an averaging artifact into the analyses produces no appreciable difference between the LLE and LSE predictions.

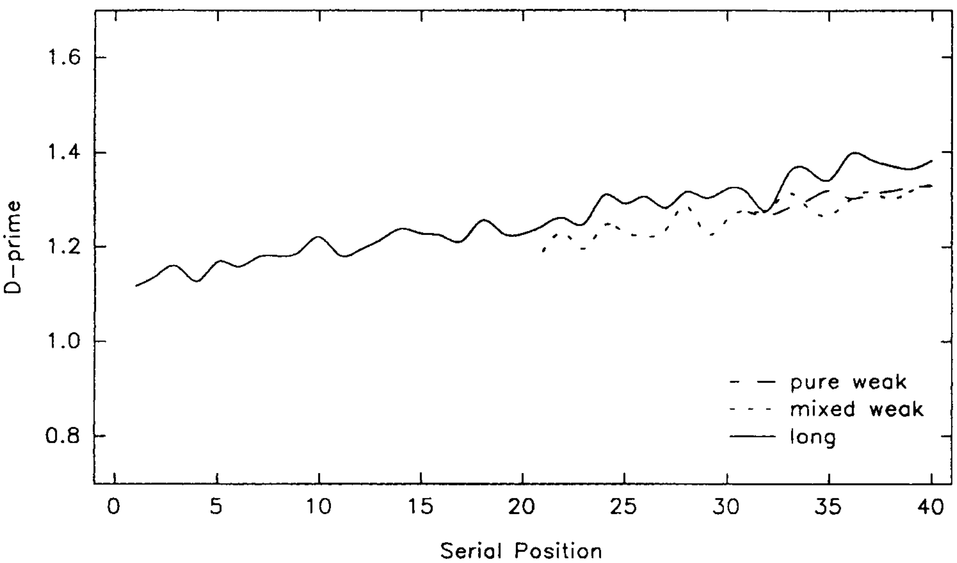

Figure 2. Predictions of recognition performance (d') for weak items and for lists of different lengths and strengths obtained by simulation of the continuous memory version of TODAM proposed by Murdock and Kahana (1993a). (The simulation uses the parameter values they suggest and is applied to the paradigm they suggest. *Pure weak* is a list of 10 singly presented items; *mixed weak* refers to a list of 5 singly presented items [the performance of which is graphed] and 5 thrice-presented items [not graphed]; *long* is a list of 40 singly presented items.)

In an analysis that we carried out but do not discuss in detail, we looked at list position in session: The slightly negative list-length outcome gradually diminished as the session continued. This effect is due to the variance: Long lists of singly presented items reduce variance (because q is greater than p and there are no repetitions to raise variance and compensate). As the session continues, the counterbalancing of lists and the introduction of many lists with repetitions reduce these differences.

For repeated items, a single measure of serial position is not clearly defined. When we summed over such items, the pure strong lists have a d' value of 3.70, and the strong items in mixed lists have a d' value of 3.77. This is again a slightly positive list-strength prediction, but it may be slightly inflated because the pure strong lists have more presentations than the mixed lists, and our analysis does not control for serial position effects. Regardless, when we analyze the data from our studies in a similar fashion, the results are in a direction opposite to the TODAM predictions.

Thus our simulation of TODAM for the case of spaced repetitions, in which we used the parameter values suggested by Murdock and Kahana (1993a), confirms the predictions from the simplified model: The continuous memory version of TODAM is unable to predict the observed data pattern, that is, a positive LLE and a missing or negative LSE.

In summary, the continuous memory version of TODAM eliminates most variance differences that are due to composition of the most recent list and as a consequence eliminates most of the predicted effects that are due to such variance differences, such as list-strength and list-length effects. These similar predictions for length and strength do not match the observed findings. In addition, those differences that are predicted (i.e., because of length) are in the wrong direction. It is true that the length misprediction can be lessened by averaging performance across all serial positions, but this does not help matters: First, in TODAM, averaging across all positions leads to the prediction of a positive LSE, for the same reasons that apply to list length. However, a positive LSE is not observed. Second, LLEs occur even when serial position and study–test lag are held constant. Thus the continuous memory version of TODAM does not predict the observed list-length and list-strength findings, which verifies the claims of Shiffrin et al. (1990).[1]

[1] The reply by Murdock and Kahana (1993b) that follows this comment tends to obscure our main point, which remains valid: Their model does not predict key elements of the pattern of data. They appear to be aware of this, although it may not be evident to the casual reader. They state that TODAM "cannot handle some of the detailed findings in the myriad LSE [list-strength effects] experiments reported by Shiffrin and his colleagues" (Murdock & Kahana, 1993b, p. 1451). Of course these detailed findings are just the ones that we have said (see Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Ratcliff, & Clark, 1990) are critical for an evaluation of TODAM. It is possible that some other version of TODAM can be generated to do the job, and we would like to see it produced. Alternatively, Murdock and Kahana argue that all empirical loopholes have not been closed and that our data pattern may not be correct: They raise the possibility that the list-length effect does not exist if lag is properly controlled. We strongly disagree, to the extent that we do not wish to devote further resources to testing the issue. If Murdock and Kahana truly believe this hypothesis, we hope they will carry out a study demonstrating their point.

References

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

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

Murdock, B. B., & Kahana, M. J. (1993b). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and Nobel (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1450–1453.

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

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

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

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214.

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

Received November 27, 1992
Revision received February 24, 1993
Accepted February 25, 1993 ■

Six Editors Appointed, 1995–2000

The Publications and Communications Board of the American Psychological Association announces the appointment of six new editors for 6-year terms beginning in 1995. As of January 1, 1994, manuscripts should be directed as follows:

- For the *Journal of Abnormal Psychology*, submit manuscripts to Milton E. Strauss, PhD, Department of Psychology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7123.

- For the *Journal of Applied Psychology*, submit manuscripts to Philip Bobko, PhD, Editor, *Journal of Applied Psychology*, P.O. Box 130, Skillman, New Jersey 08558.

- For the *Journal of Comparative Psychology*, submit manuscripts to Charles T. Snowdon, PhD, Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, Wisconsin 53706-1696.

- For the Attitudes and Social Cognition section of the *Journal of Personality and Social Psychology*, submit manuscripts to Arie W. Kruglanski, PhD, Department of Psychology, University of Maryland, College Park, Maryland 20742.

- For *Professional Psychology: Research and Practice*, submit manuscripts to Patrick H. DeLeon, PhD, JD, Editor, *Professional Psychology: Research and Practice*, APA, Room 3084, 750 First Street, NE, Washington, DC 20002-4242.

- For *Psychological Review*, submit manuscripts to Robert A. Bjork, PhD, *Psychological Review*, Department of Psychology, University of California, Los Angeles, California 90024-1563.

Manuscript submission patterns make the precise date of completion of 1994 volumes uncertain. The current editors, Susan Mineka, PhD; Neal Schmitt, PhD; Gordon G. Gallup, PhD; Abraham Tesser, PhD; Ursula Delworth, PhD; and Walter Kintsch, PhD, respectively, will receive and consider manuscripts until December 31, 1993. Should any 1994 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 1995 volumes.