Retrieval latency and “at-risk” memories

STEPHEN MADIGAN and JAMES NEUSE

University of Southern California, Los Angeles, California

and

URTE ROEBER

University of Leipzig, Leipzig, Germany

Does retrieval latency reflect variations in the strength of associations in episodic memory? In three experiments, subjects were given a single study and test trial on each of five lists of 10 paired associates. Spoken recall latencies were measured. When the subjects were later given a second test, initial recall latency was systematically related to *intertest retention*—that is, the faster the initial correct recall of a pair, the more likely a pair was to be recalled at the second test. This effect occurred at retention intervals of 5 min, 30 min, and 24 h and was present in the data for individual subjects. The results are consistent with the classical view of latency as a measure of trace strength and stand in sharp contrast with results reported by Benjamin, Bjork, and Schwartz (1998) that showed that fast retrievals from semantic memory were more poorly retained than slower ones.

In a recent analysis of curves of forgetting, Rubin and Wenzel (1997) asked the following question:

How well can one predict future forgetting? For instance, if one knows that a group of students studied a lesson until they knew it perfectly and that 1 hr later they remembered 80%, can more be said about remembering 1 week later than it will be between 0 and 80% most of the time? (p. 755)

The experiments reported here indicate that indeed more can be said about retention if recall latencies are measured at the end of the learning phase. Specifically, we show that response latencies in the first test trial (T1) of a study–test–test procedure are predictive of intertest forgetting: Relatively slow responses at T1 are much more likely to be forgotten on a later test (T2) than are faster T1 responses.

The idea that response latency reflects trace strength or associative strength is hardly new. Woodworth and Schlosberg (1954) described it this way: “It is assumed that the quicker the reaction, the stronger is the associative connection. This assumption has a sort of ‘face validity’ . . . it is also supported by the fact that in learning a new pair of associations the RT diminishes as the learning progresses” (p. 49). The latter claim has been amply documented. In multitrial paired-associates learning experiments, response latency for correct recalls decreases as percent recalled increases over learning trials (Anderson, 1995; Shapiro, 1968).

Against this background, the proposition that fast retrievals will be well retained and slow ones forgotten seems unsurprising. However, some recent as well as some older studies have suggested that the relationship between trace

strength and recall latency may not be as straightforward as the “classical” view suggests. These are studies in which response latencies for individual items have been used to predict retention for those items. Murdock (1974) presented an apparent exception to the strength–latency principle in the results of two experiments using a continuous paired-associates task with word pairs tested twice, with recall latency measured at each test. The rationale of the experiments was that “strength theory would predict that the faster the latency on the first test, the stronger the response; consequently, the higher its recall probability should be on the second test” (p. 101). No such effect was obtained. One experiment showed no consistent relationship between initial response latency (RT) and retention, and the other showed a tendency toward the opposite outcome (i.e., slower T1 responses were associated with better retention across tests). MacLeod and Nelson (1984) cited this study as one source of support for their arguments that recall accuracy and recall latency do not necessarily measure the same property of memory traces.

Some more recent research seems even more at variance with the classical strength–latency hypothesis. Replicating a finding first reported by Gardiner, Craik, and Bleasdale (1973), Benjamin, Bjork, and Schwartz (1998) showed that the time that it took to generate answers to general information questions was related to the subject’s ability to later recall the answers in a free recall task, but the relationship was the opposite of what one might have expected from the classical view of trace strength and latency. Items retrieved relatively slowly were *better* remembered on a later test than were more quickly retrieved items. Benjamin et al. hypothesized that this occurred because more “elaborated” traces were created during longer memory searches.

The experiments reported in this article had their origin in the contrast between the kinds of results just de-

Address correspondence to S. Madigan, University of Southern California, Psychology Department, Los Angeles, CA 90089-1061 (e-mail: smadigan@rcf.usc.edu).

scribed and another experimental result that does in fact appear to show precisely the latency–strength–retention relationship that Murdock (1974) argued follows from strength theory. Madigan and McCabe (1972) presented subjects with 50 lists of five paired associates with a single pair tested per list. Spoken RTs were recorded. At the end of the experimental session, the subjects were retested on all 50 pairs. Murdock's (1974) prediction was correct—that is, pairs retained from the initial to the final test had shorter RTs in the first test than did pairs recalled in the initial test but forgotten on the later test (although the situation was complicated by important effects of serial position, a point discussed in more detail later). This result appears to be the only extant support for the proposition that fast retrievals lead to good retention when measured at the level of individual items. The series of experiments reported here were intended to examine the latency–retention relationship in more detail.

METHOD

Design

The experiments were done on the basis of a simple two-stage within-subjects procedure—presentation and immediate testing (T1) of five lists of 10 paired associates each, with this test occurring immediately after list presentation; and a second test (T2) that occurred following presentation and testing of the fifth list. This test of retention occurred approximately 5 min (Experiment 1), 30 min (Experiment 2), or 24 h (Experiment 3) after the end of the study–test phase.

Subjects

The subjects were male and female undergraduates serving for course credit. There were 40 subjects in Experiment 1, 40 subjects in Experiment 2, and 32 subjects in Experiment 3.

Procedure

In all the experiments, word pairs were formed by random selection from two pools of 100 medium- to high-frequency concrete nouns (Paivio, Yuille, & Madigan, 1968), from which synonyms and rhymes had been deleted. One pool consisted of disyllabic words that were used as stimulus or cue items. The other pool consisted of monosyllables that were used as response items.

A test session began with instructions describing paired-associates learning and recall. This was followed by two practice lists of five pairs each. For each of the five lists that followed, the study trial consisted of a ready signal and then the sequence of 10 word pairs presented at 4-sec intervals. The test trial started 2 sec after the last pair was presented, with a display of a row of five asterisks that counted down at a 0.5-sec rate, followed by CRT presentation of 1 of the 10 cue words. The subjects spoke responses into a microphone and a voice key timer circuit recorded the latency. The experimenter then entered the RT and a correct or incorrect indicator on a numeric keypad. This event terminated the test period for the current pair and initiated the test for the next pair. The subjects were allowed up to 10 sec to make a response, at which point the system “timed out” and went on to the next test. There was a 5-sec pause between the last test and the start of presentation of the next list. Each subject in all three experiments had a different randomly generated set of word pairs, assigned at random to study and test positions. In order to eliminate recency effects, the test sequences were constrained so that the last two pairs in a study list never appeared in the first or second test positions.

The retention test consisted of the presentation of the cue words for all pairs that had been correctly recalled in T1, arranged in a random sequence with respect to lists and positions within lists. We did not test pairs that were not recalled in T1, because pilot work had indicated that such pairs were recalled at a very low rate in T2. The retention test events followed the same procedure as in T1. The test sequence itself consisted of all pairs that were correctly recalled in T1, presented in a random order. In Experiment 1, this testing occurred approximately 5 min after the study–test cycles. In Experiment 2, a 30-min delay was introduced between T1 and T2, during which the subject moved to a different room and performed a series of nonverbal spatial memory span tests. In Experiment 3, the subjects were dismissed following testing of the fifth list and were asked to return 24 h later.

RESULTS

Mean percentages recalled in T1 were 79.3%, 67.5%, and 75.0% in Experiments 1–3, respectively. Mean percentages recalled in T2 were 72.2%, 55.0%, and 32.7%, respectively. Conditionalized on initial correct recall, percentages retained from T1 to T2 were 90.9%, 81.5%, and 42.4% at retention intervals of 5 min, 30 min, and 24 h, respectively. Figure 1 shows the distribution of cued recall RTs for 4,117 correct recalls over all three experiments.¹ The mean RT for 387 overt recall errors over all three experiments was 2,746 msec, which is noticeably greater than the mean of 1,807 msec for correct recalls. RT distributions for cued recall are hard to find in the literature, but one recent source (Shiffrin, 1999) has presented RT data from a single-trial paired associates experiment that reproduced the main characteristics of the present data—that is, a strong positive skew, a mean and median of less than 2 sec, and error latencies with a mean of about 1 sec greater than the mean for correct responses.

The central question concerns the relationship between T1 RT and intertest retention and forgetting. Specifically, were faster responses more likely to be retained than slower responses? The answer is yes, and it shows up in several different ways of looking at the results. Figure 2 shows the percentage recalled in the second test as a function of T1 raw latency, with four classifications of latency—less than 1 sec, 1 to 1.99 sec, 2 to 2.99 sec, and equal to or greater than 3 sec. The result is clear: There was a consistent decrease in intertest retention at all three retention intervals as T1 recall latency increased.

To what extent does this result show only that “fast” subjects recall more and “slow” subjects recall less, or that “easy” pairs are retrieved more quickly than “harder” pairs? With respect to the subject differences problem, the data for all three experiments do in fact indicate that the correlations between T1 recall and median latency were strongly negative ($-.54$, $-.59$, and $-.48$ in Experiments 1–3, respectively). However, two further analyses indicate that the latency–retention relationship reflects more than covariation of recall and latency across subjects. In one analysis, we classified each T1 recall event as less than or greater than the median RT on a subject-by-subject basis and then found the percentage recalled

Figure 1. The distribution of initial correct recall latencies over all three experiments, with a mean of 1,807 msec, a median of 1,485 msec, and a standard deviation of 1,064 msec. x-axis labels are the upper limits of 500-msec bins.

in the retention test for these classifications. These results appear in Table 1 and they show that pairs recalled in T1 with less than the median latency were significantly more likely to be recalled in the retention test than were pairs recalled with greater than the median latency, an outcome that was obtained for 87 of the 112 subjects tested.

In an expanded analysis along the same lines, we used the method described by Benjamin et al. (1998) and found Vincentized proportions retained as a function of quartiles of T1 latency, with RT quartiles and proportions again computed separately for each subject. These data appear in Figure 3. The main result is that the latency-forgetting relationship in the raw RT data is present—that is, intertest retention decreased over quartiles of initial retrieval latency. In one further analysis, we found the point-biserial correlation of continuous T1 RTs and dichotomous T2 recall (1 = *recalled*, 0 = *not recalled*) for each subject. The median size of the correlation was not particularly impressive ($r_{pbis} = -.23$), but over all three experiments 81% of these correlations were negative, associating slower recall with greater intertest forgetting.²

Since each subject in each experiment had a unique set of materials, parallel analyses of pair difficulty–latency could not be conducted. In any event, we would agree with the suggestion that fast recalls and good retention will occur for “easy” pairs, but we also argue that it is the RT information that leads to the identification of such pairs in the first place. For example, consider the word pairs STATUE–KING, CARPET–BED, and STOMACH–SLEEP. They obviously have nonzero preexperimental associative strength. They were recalled in T1 with the shortest latency for each of 3 subjects, and they were also recalled successfully in the retention test. Now consider three

pairs from 3 other subjects, pairs that have much lower preexperimental strength, FLOWER–HORSE, HAMMER–LAKE, and SCISSORS–QUART. They can also be classified as “easy” pairs because they were recalled with the shortest latency for these subjects (the raw latencies were very similar to those for the easy pairs), and they were also recalled at T2. Retrieval latencies provide important information about associative trace strength.

DISCUSSION

A simple summary of the results is that “not all retrievals are created equal,” to borrow a phrase from Benjamin and Bjork (1996, p. 328). This adds an important detail to the principle that a test trial and a successful retrieval promote retention (Bjork, 1975). In particular, our results indicate that memories retrieved relatively slowly are at risk for subsequent forgetting and raise the possibility that retrievals that exceed some critical latency might not in fact produce better recall of initially untested pairs.

The empirical generalization that we have arrived at, that there is a negative correlation of latency and retention, is the opposite of the conclusion reached by Benjamin et al. (1998), that “more involved, difficult retrievals are more effective learning events than are very fluent [relatively fast] retrievals” (p. 59). Why are the two sets of results so different? The theoretically interesting possibility is that the difference between our results and those of Benjamin et al. is a consequence of real differences in the properties of episodic and semantic memory. However, the list of possibly important procedural differences between the two sets of studies is very long. In our exper-

Figure 2. Percentage of pairs retained as a function of initial recall latencies at three retention intervals.

iments, we used cued recall of associative and item information from episodic memory, whereas Benjamin et al. used a test of retrieval from semantic memory that was followed by a test of episodic free recall. In our experiments, we also controlled posttest processing by terminating a test as soon as a response was made; the Benjamin et al. procedure required subjects to make a judgment about the later memorability of the response that had just been made.³

There are some problematic aspects of the present studies that need to be acknowledged. First, what we have demonstrated is a correlation of RT and retention. We have argued that this is not a simple artifact of subject ability or item difficulty, but the correlational nature of the data remains.⁴ It is at least logically possible that the superior retention of quickly retrieved items is a consequence of quick retrieval; that is, it is a test-trial event and outcome, and not a reflection of trace properties prior to retrieval.

There are also some clear boundary conditions to the fast retrieval–good retention principle. One is that it seems to apply only to material in memory whose initial retrieval is not influenced by short-term recency effects. In a paired-associates probe procedure, immediate testing of the last two pairs in a list produces much shorter recall latencies and higher levels of immediate recall than

do pairs from earlier in the list, but these pairs are also very poorly recalled in a later retention test (the *negative recency effect*: see Benjamin et al., 1998, Experiment 3; Madigan & McCabe, 1972). In the Madigan and McCabe experiment, pairs from List Positions 1, 2, and 3 produced the same pattern of results found in the present experiments—that is, there was more intertest forgetting for slower T1 recalls. Pairs from the last two list positions showed no such effect and possibly even reversals.

Finally, although we have stated that the results follow from a continuous-strength view of associative traces, this leaves a number of questions unanswered. In particular, we do not know what the components of the RT distribution shown in Figure 1 might be. Presumably, the observed RTs are the end result of cue word encoding, associative retrieval, and response production and decision processes. The underlying distribution of associative strengths, the nature of “associative strength” itself,

Table 1
Percentage of Pairs Forgotten for Responses Less Than the Median RT and for Responses Greater Than the Median RT

	< median	> median	<i>SE</i> _{diff}	<i>t</i>	<i>p</i>
Experiment 1	7.29	12.30	1.60	3.12	.003
Experiment 2	13.74	27.03	2.57	5.00	.001
Experiment 3	48.58	65.53	2.55	5.87	.001

Figure 3. Percentage of pairs retained as a function of subject quartiles of initial recall latencies at three retention intervals.

and the function that transfers strengths into latencies are all unknown. Developing a detailed understanding of the process might be even more difficult than it has been for the simpler case of recognition memory latencies (Murdoch, 1985).

These issues notwithstanding, our results show a clear and consistent negative correlation between initial retrieval latency and later recall. The prediction of retention from latency holds at short (5 min) as well as at much longer (24 h) retention intervals. Recall latency seems to reflect some property of associative memory traces that is related to the durability of those traces, and it does so in a way that is consistent with the time-honored characterization of RT as a measure of associative strength—that is, fast retrievals from episodic memory are associated with better retention than are slower retrievals.

REFERENCES

- ANDERSON, J. R. (1995). *Learning and memory*. New York: Wiley.
- BENJAMIN, A., & BJORK, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition: The 27th Carnegie Symposium on Cognition* (pp. 309-338). Hillsdale, NJ: Erlbaum.
- BENJAMIN, A., BJORK, R. A., & SCHWARTZ, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, **127**, 55-68.
- BJORK, R. A. (1975). Retrieval as a memory modifier. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 124-144). Hillsdale, NJ: Erlbaum.
- BREWER, J. A., ZHOU, Z., DESMOND, J. E., GLOVER, G. H., & GABRI-ELI, J. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, **281**, 1185-1187.
- GARDINER, J. M., CRAIK, F. I. M., & BLEASDALE, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, **1**, 213-216.
- MACLEOD, C. M., & NELSON, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica*, **57**, 215-235.
- MADIGAN, S., & MCCABE, L. (1972). Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **10**, 101-106.
- MURDOCK, B. B., JR. (1974). *Human memory: Data and theory*. Potomac, MD: Erlbaum.
- MURDOCK, B. B., JR. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, **13**, 511-521.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, **76**(1, Pt. 2), 1-25.
- RATCLIFF, R., & MURDOCK, B. B., JR. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.
- ROHRER, D., & WIXTED, J. T. (1994). An analysis of latency and inter-response time in free recall. *Memory & Cognition*, **22**, 511-524.
- RUBIN, O., & WENZEL, O. (1997). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, **103**, 734-760.
- SHAPIRO, S. I. (1968). Paired-associate response latency as a function of free association strength. *Journal of Experimental Psychology*, **77**, 223-231.
- SHIFFRIN, R. M. (1999). 30 years of memory. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model* (pp. 17-33). Mahwah, NJ: Erlbaum.
- WAGNER, A. D., SCHACTER, D. L., ROTE, M., KOUTSTAAL, W., MARIL, A., DALE, A. M., ROSEN, B. R., & BUCKNER, R. L. (1998). Building memories: Remembering and forgetting of verbal experience as predicted by brain activity. *Science*, **281**, 1188-1191.

WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology*. New York: Holt.

NOTES

1. The smooth curve was produced by a least-squares fit to an ex-Gaussian distribution. This distribution has been used widely to describe response latencies in recognition (Ratcliff & Murdock, 1976) and free recall (Rohrer & Wixted, 1994). It obviously produces an excellent fit for cued recall as well (variance accounted for is 99%). We thank John Wixted for making the curve-fitting software available to us.

2. Following a reviewer's suggestion, we also computed the Goodman-Kruskal gamma for each subject, a procedure that discards all ties on either or both variables. The results with this measure were essentially the same as with r_{pbis} —that is, the median value of gamma was $-.25$, and 83% of these correlations were negative.

3. Benjamin et al. (1998) do not provide any details about the duration of the initial retrieval trials (e.g., maximum response time allowed and time allowed for memorability judgments).

4. This point might be kept in mind when one considers the results of two recent studies of brain imaging and memory (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Wagner et al., 1998). Using the same kind of correlational design as in the present studies, these authors concluded that levels of cortical activation produced at the time of presentation or encoding were predictive of subsequent remembering (although they actually reported cortical activity levels as a function of retention test performance).

(Manuscript received December 28, 1998;
revision accepted for publication July 2, 1999.)