

A comparison of recognition and savings as retrieval measures: A reexamination

LINDA KNAPP GRONINGER and LOWELL D. GRONINGER
University of Maryland Baltimore County, Baltimore, Maryland 21228

The present study was designed to extend the findings of Nelson (1978) in determining the relative sensitivities of recognition and savings for detecting small amounts of information in memory. Subjects learned a list of 20 number-noun paired associates to a recognition criterion. After a retention interval of 2, 3, or 4 weeks, subjects returned for tests of recall and recognition. Then, at relearning, some of the nonretrieved items remained the same as at original learning, and some new items were created by a re-pairing of numbers and nouns. After one relearning study trial, retention tests were again given. Statistical tests showed significant savings effects on all dependent measures. It was concluded that savings is a more sensitive retrieval measure than recognition and that subthreshold memory traces can exist and can be incremented by the process of concatenation. It was suggested that a reinstatement of old mediators may also play a role in the savings effect.

Comparisons of retrieval measures have provided many insights and much empirical data with respect to the understanding of human memory. For example, comparisons of recall and recognition have been the focal point for investigating one- vs. two-process theories of memory, storage vs. retrieval issues, and the principle of encoding specificity. To date, however, memory theorists have largely overlooked the potential theoretical contribution that might be made from an analysis of the savings method, one of the first techniques systematically used to measure retrieval (Ebbinghaus, 1885/1964).

A likely reason for the dearth of studies investigating savings is the difficulty of making direct comparisons between savings and other retrieval measures. As Crowder (1976, p. 371) points out, "How do we compare a 50% savings in relearning trials to six items correctly recalled from a list of ten?" Only recently has a technique been developed by Nelson (1971) that allows direct comparisons between savings and recognition while eliminating many of the methodological problems involved in measuring savings through relearning trials (see Bahrick, 1967; Nelson, 1971, 1978, for a review of these problems). Nelson's (1971) revised savings paradigm can be described as follows.

In the original learning (OL) session, using a study-test procedure, subjects learn a list of number-noun paired associates. No mention is made of any further retention test. After a retention interval, subjects return

This research was part of a master's thesis submitted by the first author to the University of Maryland Baltimore County. The project was aided considerably by the advice and support of the committee members: Herbert Weingartner, Mark Moody, and Marilyn Wang. Computer time was provided by the Computer Science Center of the University of Maryland. Requests for reprints should be sent to Lowell D. Groninger, University of Maryland Baltimore County, Catonsville, Maryland 21228.

for a second session. Memory for the previously learned associations is assessed using a self-paced cued recall retention test. During relearning (RL)¹ items missed on the retention test are either (1) re-presented (old items) or (2) changed by randomly re-pairing the cues and targets of items missed previously (new items). The re-pairing assures that if there is any forgetting of responses as opposed to forgetting of associations, there will be no differential effects on the same vs. changed conditions during RL. Savings is then measured as a difference between the relearning of nonrecalled old items and the learning of new items.

Nelson's (1971) revised savings paradigm deals with a residual trace that, by definition, is weaker than that trace required for the recall of a target item. Of particular relevance to the present study is the relationship between savings and recognition. Traditionally, studies (e.g., Luh, 1922; Postman & Rau, 1957) have suggested that recognition is a more sensitive measure of retrieval than savings. However, Nelson (1978) showed that the reverse can be true when he demonstrated savings for items that had been both nonrecalled and nonrecognized on the retention tests.

A comparison of savings and recognition is important in determining the nature of memory traces. For example, if savings were a more sensitive measure of retrieval than recognition, it would imply that some subthreshold or residual traces of memory exist that are not retrievable. If the reverse were true (i.e., that recognition is a more sensitive measure of retrieval than savings), it would imply that a memory trace exists at or above threshold or not at all. That is to say, there would be an all-or-none trace in the sense that if it existed at all it could be detected by a recognition test.

Although Nelson's (1971) paradigm has dealt with many of the methodological problems that were present

in previous savings research, there remains a possible bias in his procedure with respect to a comparison of recognition and savings. There is evidence (Carey & Lockhart, 1973; Hall, Grossman, & Elwood, 1976; Jacoby, 1973; Tversky, 1973) that subjects' anticipation of the mode of the expected memory test influences the manner in which the test items are encoded. Of particular relevance to the present study is an experiment by Jacoby (1973), in which he presented four categorized lists for a single study-test trial. For the first three lists, test mode (free recall, cued recall, or recognition) was held constant. The mode of testing on the fourth list was then factorially combined with the test mode of the first three lists. The results showed cued recall performance to be higher when subjects anticipated cued recall rather than free recall or recognition. That is, when there was congruity between study and test modes, performance improved.

In Nelson's (1978) study, acquisition was measured via cued recall. Since savings was also measured by cued recall, there was congruity between the study mode at OL and the savings test mode. However, during the retention tests, when recall and recognition were used to classify items as forgotten or not, the most sensitive of these tests, recognition, was incongruous with the study mode at OL (recall). Thus, savings was being measured by the same method as OL (recall), but it was being compared with recognition memory, which had been evaluated using a method (recognition) incongruous with that used at OL. Therefore, there existed a bias in favor of demonstrating that savings is a more sensitive measure than recognition.

The present study was designed to address this issue. An incongruous study-test relationship was established in that acquisition was measured via recognition.² Items were then classified as forgotten or not on the basis of recognition and recall tests, and savings was measured via a cued recall procedure. If it can be shown that savings is a more sensitive measure of retrieval than recognition under circumstances that are biased against savings and in favor of recognition, then Nelson's (1978) conclusion that savings is a more sensitive measure than recognition for detecting small amounts of information in memory will be unequivocally supported. If, instead, savings is not found to be more sensitive than recognition, some question will remain as to the generality of Nelson's conclusion.

METHOD

Materials

The list for a given subject was composed of 20 paired associates. The cues were two-digit numbers of low (1.54-1.99) association value from the Battig and Spera (1962) norms. The targets were four- to six-letter nouns with high (means greater than 6 and standard deviations less than 2.25) imagery, concreteness, and meaningfulness ratings on the Paivio, Yuille, and Madigan (1968) norms. The nouns were also of the highest (AA) frequency value on the Thorndike-Lorge (1944) norms. The 400 cue-target combinations, resulting from crossing the 20 cues with the 20 targets, were randomly divided into 20 lists with the con-

straint that each number and each noun appear once in each list. Every list was used once before any list was used again, and no list was used more than twice. Items were printed in black ink on 3 x 5 in. index cards. A hand-held stopwatch was used for timing. The cards were presented manually by the experimenter, and oral responses were given by subjects.

Acquisition

A list of four letter-letter pairs (e.g., M-T) was first learned by subjects to familiarize them with the task. Using the procedure and timing outlined below for OL, this practice task continued until the subject could recognize each pair once when each was presented among four distractors (e.g., M-R, M-L, M-B, M-T, M-X).

Following this task, the OL task was given. The number-noun pairs were presented at a 5-sec rate. After all 20 items had been presented, subjects read numbers for 20 sec to eliminate recall from short-term memory. Following this digit-reading task, subjects were given a recognition memory task, in which the number cue for each item was paired with its noun target and four distractors, each of which had been targets for other number cues at OL. For example, if the OL item was 17-QUEEN, the items on the recognition memory task might be 17-RIVER, 17-QUEEN, 17-GARDEN, 17-FIRE, 17-BOOK. The OL item and the four distractors were randomly ordered for this task. Subjects were allowed 8 sec to say aloud which noun had been paired with that number during study. They were then asked to rate their confidence in the correctness of their choice on a 3-point scale from 1 ("very sure") to 3 ("just guessing"). If a response was correct but given a confidence rating of 3, it was not considered to be learned. After each test trial was completed, there was a brief delay while the experimenter rearranged the cards for restudy. To minimize overlearning, once a given item was correctly recognized, it was eliminated from the study and test phases of the list. Thus, the criterion for acquisition was one correct response per item. After OL was completed, to allay any suspicion of a subsequent retention test, subjects were given a relationship-reasoning test, ostensibly the second test in a battery of four. Subjects were asked to return after a specified retention interval to complete the battery.

Retention Tests and Relearning

Subjects returned for the RL session 2, 3, or 4 weeks after OL. First, a self-paced cued recall retention test was given, in which each number cue was shown for as long as was necessary for the subject to respond. Subjects were encouraged to guess but were not forced to respond. After all 20 items had been tested for recall, the procedure was repeated, since occasionally items that are incorrect on a first test trial are correct on a second one. The repetition served to define a more accurate pool of nonrecallable items. Next, a forced-choice recognition test was administered. Subjects were given a sheet on which 20 cues and 20 targets were listed, and they were asked to match cues and targets in order to recreate the list of items they had studied at OL. Using this procedure, the recognition test could not serve as a study trial for old items. Therefore, the possibility of any differential effects for old and new items at RL arising from non-equivalent exposure was eliminated. For example, had the target been paired with its cue along with a small number of distractors, the physical presentation of the correct pair conceivably might have increased subthreshold strength and thereby served as a study trial. Additionally, an increment in learning might have occurred if subjects were able to eliminate, as incorrect, some or all of the distractors. While the experimenter selected the items for relearning, subjects worked on a puzzle.

To construct the RL list, the original items were divided into three pools: (1) incorrectly recalled and incorrectly recognized, (2) incorrectly recalled but correctly recognized, and (3) correctly recalled. For each of the pools, approximately one-half of the items remained the same as at OL (old items) and approximately one-half were changed by the random re-pairing of the cues and targets of the remaining items in a given pool. For example, from

a pool of 44-BOOK, 17-DOCTOR, 96-CHAIR, and 95-CAMP, if the old items were 17-DOCTOR and 95-CAMP, the new items would be 96-BOOK and 44-CHAIR. Since the focus was on memory for associations, this procedure was used in order to eliminate any differential effects at RL resulting from learning the identity of target words (as opposed to learning the associative links). Similarly, to insure that remembered associations would not differentially interfere with the relearning of new associations that came from forgotten items, the cues and targets were never crossed between the three pools listed above. For example, if 40-SHIP had been correctly recalled on the retention test and 26-MONEY had been neither correctly recalled nor correctly recognized, the learning of a new item 26-SHIP could be affected by the memory of SHIP being paired with 40 at OL. (Time constraints did not permit the elimination of the possibility that incorrect choices on the recognition test might appear as new items at RL. However, post hoc analysis showed that this occurred only 1.7% of the time. Further, these items were not recalled with greater frequency than other new items, so it can be concluded that no bias was present from this procedure.)

Immediately prior to RL, subjects were told that some items would be old and some new. The 20 RL items were randomly ordered, and the same study and number-reading procedures were used as in OL. After the digit reading had ended, three measures of retrieval were obtained. First, a cued recall retention test for the RL items was given, in which subjects had 8 sec to respond with the correct noun. Next, subjects were given a self-paced cued recall retention test for the RL items in order to obtain a more sensitive measure of retention. Finally, a second form of the forced-choice recognition test described above was administered. Subjects were then debriefed and dismissed.

Subjects

The subjects were 35 introductory psychology students from the University of Maryland Baltimore County who were chosen without regard to gender and who participated as part of a course requirement. Originally, 12 subjects were randomly assigned to each retention interval, but due to attrition and the need for 1 subject to be rescheduled, at RL there were 11 subjects each in the 4- and 2-week retention groups and 13 in the 3-week retention group.

RESULTS

The percentages of items incorrectly recalled on the retention tests after the 2-, 3-, and 4-week retention intervals were 81.8%, 81.15%, and 86.36%, respectively. Retention performance for recognition was somewhat higher than that of recall. In the 2-week retention group 15% of the nonrecalled items were correctly recognized, whereas in the 3-week retention group 16.1% of the incorrectly recalled items were recognized, and in the 4-week retention group 12.1% of the nonrecalled items were recognized.

In comparing the data for the three retention intervals, in terms of both the percentages of items incorrectly recalled on the retention tests and the percentages of savings demonstrated, it was apparent that the delay intervals used did not produce meaningful differences in either the retention of items or in the pattern of savings scores. Therefore, for the primary analyses, the data from the three delay intervals were combined in order to increase statistical power.

Orthogonal binomial tests of the difference between two proportions (McNemar, 1969) were used to determine the degree of difference between the relearning

performance of old and new items. Across the three dependent measures, relearning of old items was consistently better than learning of new items, thus demonstrating a savings effect. First, consider the relearning performance of items that had been correctly recognized after being incorrectly recalled on the retention test. Statistical tests (one-tailed) on the savings scores in Table 1 showed a significant savings for the 8-sec cued recall test ($z = 3.016$, $p < .005$), the self-paced cued recall test ($z = 2.203$, $p < .05$), and the recognition test ($z = 1.764$, $p < .05$). Second, consider the relearning performance of items that had been neither recalled nor recognized on the retention tests. Again, savings was demonstrated on all three dependent measures: the 8-sec cued recall test ($z = 3.125$, $p < .001$), the self-paced cued recall test ($z = 1.64$, $p = .05$), and the recognition test ($z = 2.808$, $p < .05$).

Since the data in the above analyses were pooled across subjects, additional analyses were done on an individual subject basis as a check on the consistency of the findings. For those items neither recalled nor recognized on the retention tests, z scores were calculated for each subject using the same formula as for the pooled data. Using t tests it was determined that the average of the z scores was significantly different from zero across the three dependent measures, again demonstrating significant savings [for the 8-sec cued recall data, $t(34) = 3.45$, $p < .01$; for the self-paced cued recall data, $t(34) = 2.32$, $p < .05$; and for the recognition test data, $t(34) = 4.15$, $p < .001$]. These tests were not performed for the non-recalled but recognized items because the proportion of subjects who had both old and new items in this category was not large enough to warrant such analysis.

In summary, the results show that significant savings occurred at RL on all three dependent measures for items that were neither recalled nor recognized and for items that were incorrectly recalled but correctly recognized on the retention tests.

DISCUSSION

Based on previous findings concerning study-test relationships (Jacoby, 1973), some question about the occurrence of a savings effect for nonrecognized items (Nelson, 1978) can be raised. The present study was designed to answer these questions. Savings was found in the present study when items were classified as nonrecognized by a method congruous with OL study mode (recognition), while savings was measured by an incongruous method (recall). Therefore, the results of the present study lend unequivocal support to Nelson's (1978) conclusion that savings is a more sensitive measure than recognition for detecting small amounts of information in memory.

This demonstration of the savings phenomenon, while providing information about memory processes, also raises several theoretical issues. For example, if savings is a more sensitive measure of retrieval than recognition, the implication is that a trace exists that is of insufficient strength to allow an item to be recognized. This suggests that the memory trace contains a subthreshold component and that the trace is brought to threshold, that is, made accessible to memory retrieval, by repeated exposure. A corollary of this position is that a memory trace can be established incrementally at a subthreshold level (a process that Nelson, 1978, refers to as concatenation).

Table 1
Percent Correct and Savings for Nonrecalled Items at Relearning

Relearning Study Condition	N	Cued Recall		Recognition
		8 Sec	Self-Paced	
Incorrect Recognition Test Performance				
Old	248	38.3	38.3	46.0
New	249	25.3	31.3	33.7
Percent Savings		13.0	7.0	12.3
Correct Recognition Test Performance				
Old	40	60.0	65.0	62.5
New	44	27.3	40.9	43.2
Percent Savings		32.7	24.1	19.3

If some assumptions are made about memory traces, support for a concatenation position can be found in Table 1, in which the magnitude of the savings effect is shown to be greater for the not recalled but recognized items than for the not recalled, not recognized items. First, assume that memory traces are stronger in the not recalled but recognized category than in the not recalled, not recognized group. Second, assume that a relearning trial adds an increment to the memory traces. Since the traces are stronger for the not recalled but recognized items, the increment in strength resulting from the relearning trial would bring more items above threshold, thereby producing the greater savings effects.

Although this explanation is useful, it is basically descriptive and does not address the issue of the mechanism by which the traces are strengthened and brought to or above threshold. One suggestion concerning the underlying processes involved in savings is that exposure to an item at RL helps to reinstate that item's retrieval context, which had been established during OL but had been forgotten during the retention interval. For example, for the item 49-MONEY, a mediator of "the gold rush of 1849" might be established to aid in the recall of the target MONEY. During the retention tests the link between the cue, 49, and "the gold rush of 1849" might not have been strong enough for the mediator to be accessed, even during recognition. However, when 49 was again paired with MONEY at RL, it is possible that the original mediating context was reestablished and that recall and/or recognition was thereby facilitated.

The reasoning behind this explanation is as follows. First, assume that some mediators are relatively low in a mediator response hierarchy so that they are not likely to be regenerated from the cue alone. Thus, the mediator is not likely to be available to aid recall after a retention period of several weeks. Next assume that in recognition tests the distractors create a situation that is not conducive to extensive searches for mediators because of either self-imposed time constraints or high interference due to the presence of distractor items. Under these conditions the mediator may not be generated. However, during the re-pairing of the items at RL, a direct attempt to find a mediator may be made, since conditions are similar (e.g., the cues and targets are presented together with instruction to learn) to those in which the mediator was originally generated. Thus, it is at RL that the mediator is most likely to be reproduced. The feasibility of this contextual explanation is supported by the studies of Smith (1979) and Smith, Glenberg, and Bjork (1978), which showed strong context effects on recall from environmental sources. Smith (1979) also found improvement in recall

when subjects imagined themselves in the original learning situation. It seems but a small step from the situational cues involving the environment in these studies to the situational cues involving tasks in the present study.

REFERENCES

- BAHRICK, M. P. Relearning and the measurement of retention. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 89-94.
- BATTIG, W. F., & SPERA, A. Rated association values of numbers from 0-100. *Journal of Verbal Learning and Verbal Behavior*, 1962, 1, 200-202.
- CAREY, S. T., & LOCKHART, R. S. Encoding differences in recognition and recall. *Memory & Cognition*, 1973, 1, 297-300.
- CROWDER, R. G. *Principles of learning and memory*. New York: Wiley, 1976.
- EBBINGHAUS, H. [*Memory: A contribution to experimental psychology*] (H. A. Ruger & E. R. Bussenius, trans.). New York: Dover, 1964. (Originally published, 1885.)
- HALL, J. W., GROSSMAN, L. R., & ELWOOD, K. D. Differences in encoding for free recall vs. recognition. *Memory & Cognition*, 1976, 4, 507-513.
- JACOBY, L. L. Test appropriate strategies in retention of categorized lists. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 675-683.
- LUH, C. W. The conditions of retention. *Psychological Monographs*, 1922, 31(3, Whole No. 142).
- MCNEMAR, Q. *Psychological statistics* (4th ed.). New York: Wiley, 1969.
- NELSON, T. O. Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, 10, 568-576.
- NELSON, T. O. Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 453-468.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph*, 1968, 76(1, Pt. 2).
- POSTMAN, L., & RAU, L. Retention as a function of the method of measurement. *University of California Publications in Psychology*, 1957, 8, 217-270.
- SMITH, S. M. Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, 5, 460-471.
- SMITH, S. M., GLENBERG, A. M., & BJORK, R. A. Environmental context and human memory. *Memory & Cognition*, 1978, 6, 342-353.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College, Bureau of Publications, 1944.
- TVERSKY, B. Encoding processes in recognition and recall. *Cognitive Psychology*, 1973, 5, 275-287.

NOTES

1. The symbol RL will be used throughout this paper to designate the relearning stage of the savings paradigm. When the process of relearning is referred to, the term will be spelled out.

2. Nelson (1978) suggested a similar manipulation as a way of adding generality to his results.

(Received for publication February 4, 1980.)