Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 2, 195–212

0278-7393/88/$00.75

Forgotten but Not Gone: Savings for Pictures and Words in Long-Term Memory

Colin M. MacLeod
University of Toronto, Scarborough, Ontario, Canada

Five experiments examined the relearning of words, simple line-drawing pictures, and complex photographic pictures after retention intervals of 1 to 10 weeks. For those items that were neither recalled nor recognized, the identical item was relearned better than an unrelated control item, as measured by a recall test following relearning. This relearning advantage in recall held for all three classes of material and extended to the cross-modality case (i.e., picture–word and word–picture) and the same-referent case (i.e., two pictures of the same object). However, recognition tests of relearning failed to detect this same relearning advantage for apparently forgotten items. Taken together, these findings conflict with the existing account of savings. Most fundamental, the classic argument that relearning serves a trace-strengthening function is undermined by the observed recall-recognition contrast. An alternative explanation of savings is suggested wherein relearning assists retrieval of information, thereby affecting recall in particular.

We all know stories about people who believe they have entirely forgotten some event or skill, and then readily recover it when given the opportunity to learn it again. A woman spoke French as a young child in pre-war Belgium but has had no exposure to the language since moving to the United States and claims no familiarity with the language as an adult. Over 40 years later, she visits Paris and finds to her surprise that she picks up the language with apparent ease. A man memorized Wordsworth's "I Wandered Lonely as a Cloud" in elementary school but has no recollection of having done so when his child shows the poem to him 30 years later. In helping the child to memorize the poem, however, the father learns it again quickly, perhaps never even recognizing why.

Such anecdotes abound. The general conclusion is that *relearning* is faster and easier than original learning. Ebbinghaus (1885) called this relearning advantage *savings*, presumably because the partial information still "saved" in memory was what assisted learning on the next encounter. Interestingly, in line with the foregoing illustrations, this savings advantage can even occur without awareness many years after the original learning (Burtt, 1932, 1941; Titchener, 1923). Without question, the savings effect is a powerful one. Undoubtedly, it is also a much more pervasive phenomenon than we might suspect from the anecdotes.

Now a century old, relearning has never gained the popularity of recall and recognition as measures of retention in memory research. Although relearning was the preferred method to study memory for Ebbinghaus (1885), few investigators have followed his lead. Until recently, the handful of studies using relearning typically had done so to contrast different measures of retention (e.g., Bahrick & Bahrick, 1964; Luh, 1922; Postman & Rau, 1957), not to study relearning in its own right. However, in the past 15 years, the studies of Nelson and his coworkers (Nelson, 1971, 1978; Nelson & Rothbart, 1972; Nelson, Fehling, & Moore-Glascock, 1979; Nelson, Gerler, & Narens, 1984) and others (Conover & Brown, 1977; Groninger & Groninger, 1980; MacLeod, 1976) have demonstrated the value of exploring savings in the traditional list-learning framework. Other studies have extended the study of savings to a broader domain, including learning new skills (e.g., Kolers, 1976). Savings is a very sensitive measure of memory—perhaps the most sensitive (Nelson, 1978)—and represents an excellent tool for the study of otherwise inaccessible memories, especially when a long retention interval is involved. Savings research can tell us much that might not be uncovered in other types of memory studies.

The Savings Paradigm

For his classic studies of memory, Ebbinghaus (1885) used a relearning procedure. He first learned each list of nonsense syllables to a criterion of error-free serial recitation. He then relearned the list to the same errorless criterion on a second occasion. Savings was indexed by the advantage of relearning over original learning in terms of number of trials to criterion. The lower was the ratio of relearning trials to original learning trials, the greater was the savings. It was from relearning/savings data that he generated his famous forgetting curve that appears in virtually every introductory psychology text.

This research was supported by Grants A7459 and E6532 from the Natural Sciences and Engineering Research Council of Canada and by funds provided by the University of Toronto. The original article was prepared while the author was on study leave at the University of Queensland.

For collecting the data, my thanks go to Deborah Paes (Experiment 1), Geoffrey Lowrie (Experiment 2), and P. J. Lootsteen (Experiments 3, 4, and 5). I also thank Barbara Tversky and Endel Tulving for providing the materials for Experiments 3 and 4, respectively. I am grateful to Bert Forrin, Sam Glucksberg, Tom Nelson, Henry Roediger, Dan Schacter, Marilyn Smith, Gay Snodgrass, the members of Endel Tulving's research group, and an anonymous reviewer for their helpful criticism of earlier drafts.

Correspondence concerning this article should be addressed to Colin M. MacLeod, Division of Life Sciences, University of Toronto, Scarborough Campus, Scarborough, Ontario, Canada M1C 1A4.

Occasionally, investigators have attempted to deal with some of the perceived flaws in the Ebbinghaus technique (see, e.g., Bunch, 1941; Bahrick, 1967), but no full-scale revision was attempted until Nelson's work (1971; see also 1985). Because Nelson et al. (1979) present a detailed discussion of the procedure, only a general outline will be provided here. In the acquisition session (Session 1), the subject learns a list of paired associates with numbers as stimuli and words as responses (e.g., 27—CHAIR, 56—SNAKE). Using paired associates permits investigation at the level of the single item. The typical list length is 20 items, and the acquisition criterion is fairly stringent (from once correct per item to the entire set correct in a single test). Subjects first study the entire list and then are tested on it, with a 5–8 s/item rate for both study and test. The test requires recall of the correct word in the presence of just the number. Study-test cycles are repeated until the subject attains criterion, at which point the subject is dismissed without mention of the second session.

Session 2 occurs several weeks or months later, to permit sufficient forgetting. The first stage is a self-paced recall test for all items, to discriminate those remembered from those forgotten. Then, there is a single study trial on the relearning list. In its simplest form, this list is made up of two kinds of pairs, those identical to original pairs (*same*) and those unrelated to original pairs (*changed*). In the foregoing examples, a same pair might be 27—CHAIR, whereas a changed pair might be 56—ASHTRAY.[1] The contrast between same and changed items for remembered pairs simply indicates whether the set of items used shows normal transfer effects and is therefore a suitable set for studying savings. Of central interest are the forgotten pairs. Here, superior performance in the same condition over that in the changed condition is taken to indicate that some information has been saved in the memory traces of the pairs, despite the subject's initial inability to recall them.

The major change Nelson made in Ebbinghaus's procedure concerned how savings was measured. For Ebbinghaus, this was based on the number of relearning trials to reach criterion again on the entire list. For Nelson, savings was reflected in the advantage of same items over changed items on a single relearning trial, but only for forgotten items. Nelson argued that his was a better measure because differences in relearning are largest on the first trial and because only relearning of the forgotten items provides information about the nature of savings. In this setting, trials to criterion is a relatively insensitive measure, and the serial nature of the Ebbinghaus procedure complicates the study of savings.

The Pattern of Savings

Nelson's research program has produced several consistent findings. First, even when unrecalled (Nelson, 1971) or unrecognized (Nelson, 1978) after the retention interval, an identical word is always relearned better than an unrelated control word is learned. This indicates that something is saved in the memory trace of a forgotten item. Studies of various original learning to relearning relations other than identity help to specify what sorts of information make up this savings residue. Nelson and Rothbart (1972) showed that part of this residue is acoustic because homophones show savings (e.g., when TAX replaces TACKS). MacLeod (1976, Experiment 1) showed that part of the residue is semantic because bilinguals show savings for translation equivalents (e.g., when CHEVAL replaces HORSE for French–English bilinguals).

The major study thus far is that by Nelson et al. (1979), investigating what sorts of semantic information are saved in the trace of a forgotten item. Both superordinate information and subordinate information are saved, so that VEHICLE and BUICK are relearned better than a control word like PRINCE even though the original item, CAR, apparently was forgotten. However, associates (e.g., TRUCK) do not show savings, nor do antonyms (e.g., for the original item LOVE, the word HATE does not show savings). Perhaps most surprising, synonyms do not show savings over several experiments—for the forgotten item CAR, AUTO showed no relearning advantage over PRINCE.

The Savings Account

The relearning picture is considerably more complicated today than it was for Ebbinghaus. As Nelson et al. (1979, p. 241) pointed out, there is no theory that predicts their pattern of savings. How, then, is one to understand savings? Nelson has offered two main ideas. First, and most central, he saw relearning as a trace-strengthening operation, much as Ebbinghaus did. In his words, "the subthreshold memory strength from a nonrecognized item might be concatenated with an extra amount of memory strength from the relearning study trial (in the case of an old item), thereby producing an overall amount of memory strength that is sufficient for recall" (Nelson, 1978, pp. 466–467). Emphasis clearly is placed on the information stored in the trace, which is seen as helping a relearned item exceed a performance threshold. This is an intuitively appealing idea.

Nelson's second idea is less general, intended as it is to explain the unusual pattern of semantic savings just described. As Nelson et al. (1979, p. 242) put it, "semantic information about nonrecalled items is saved not at the same level of

---

[1] Here, the stimulus set is constant, but changed pairs have entirely new responses (an AB-AC design). It could be argued that any residue from the original pair could actually interfere with the learning of such a changed pair. In keeping with such an argument, entirely new pairs—both stimulus and response—would constitute the appropriate control (an AB-CD design where 95—PUMPKIN would appear only in the relearning list). Both procedures have been used in previous savings work (cf. MacLeod, 1976; Nelson, Fehling, & Moore-Glascock, 1979).

Of course, the expectation is that estimates of savings would be smaller using the entirely new pairs (AB-CD) as the baseline because such pairs will not suffer interference (to the extent that any interference arises) from their relation to original learning pairs. Thus, such a control may be seen as providing a conservative estimate of savings. On the other hand, if the interest is in whether savings can be observed under certain conditions, the same-stimulus pairs (AB-AC) should provide the best chance because the expectation is that they will provide a lower baseline to the extent that interference is operating. In fact, these two control conditions did not differ significantly in the context of a savings experiment (Muir, 1982).

inclusiveness as the originally learned item, but only at higher or lower levels of inclusiveness." Thus, associates, synonyms, and antonyms should not show savings, but superordinates and subordinates should show savings, exactly as Nelson et al. (1979) found. However, this description of the data does not handle savings for translation equivalents in bilinguals (MacLeod, 1976) unless the additional assumption is made that translation equivalents share the identical semantic information. Still, these two propositions nicely capture most of the data available at present and provide a framework for future studies of savings. One major aim of the new experiments to be presented here is to put these ideas to a further test.

The New Experiments

Thus far, the bulk of savings research has been restricted to chains of nonsense syllables and isolated words in paired-associate recall. But the procedure potentially is far more general. Compared to recall and recognition, we know relatively little about the properties of savings. The five new experiments contained herein extend relearning/savings research in four ways, the initial two following the recommendations of Nelson et al. (1979, p. 247). First, the new experiments explore savings for remembered information beyond single words; here, the focus is on pictures. Second, new relations at the same level of inclusiveness are studied, in this case a word and its corresponding picture or two different pictures having the same conceptual referent. The third extension stems from a recommendation by Nelson (1978, p. 466) to examine acquisition by recognition as well as by recall. Both are used here. Finally, although Nelson conceives of evaluating relearning only by recall, the present studies use both recall and recognition to measure degree of success in relearning. The decision to use two indices of relearning will turn out to be crucial.

All of these extensions will help to increase our understanding of relearning particularly, but also more generally of its relation to the other measures of retention. This is of more than archival value. Much recent work in memory has emphasized the role of awareness, both in normal remembering (Jacoby & Witherspoon, 1982) and in amnesia (Graf & Schacter, 1985). Relearning would seem to be an ideal tool for studying memory without awareness (cf. Nelson, 1985), but much more must be known about the tool first.

At the outset, the aim of these new experiments was a modest one: to expand the domain of savings to include pictures and to survey how relearning relates to recall and recognition. Viewed from the opposite perspective, however, these new studies will also indicate whether relearning might be put to use profitably in the study of picture memory. The picture memory literature relies on recognition test data and, to a lesser extent, on recall test data (cf. Loftus, 1982); there is no existing study of savings for pictures. This can be seen as the second main purpose of the present research. The third purpose is more theoretical: These experiments also address the trace-strengthening and the levels of inclusiveness propositions of Nelson et al. (1979). As will emerge through the course of this article, these appealing ideas appear to be insufficient to handle the observed pattern of savings in long-term memory.

Experiments 1A and 1B

The first pair of experiments set out to examine savings for words and simple pictures, thereby providing a bridge to the existing studies of savings for verbal materials. The plan was to investigate relearning of all four combinations of materials: words followed by words, words followed by pictures, pictures followed by words, and pictures followed by pictures. The logic followed that of MacLeod (1976) with the two languages of a bilingual, and permitted a further test of whether there is savings for same-level items, in this case pictures and their corresponding words. As a first step, it seemed appropriate to use simple, easily labeled pictures so that the comparison to words would be quite direct (see e.g., Snodgrass, 1984). For consistency with previous studies, Nelson's (1971) procedure was followed closely, the only change being the manipulation of the materials (as in MacLeod, 1976). Would simple pictures show savings? Would a modality shift from word to picture or vice versa affect relearning? These were the questions at which Experiment 1 was aimed.

Method

Subjects. The participants were undergraduate students in an introductory psychology class at the Scarborough Campus of the University of Toronto. They were recruited by telephone, volunteering to take part in lieu of a laboratory requirement in the course. There were two versions of the experiment. Sixteen subjects completed the 4-week version, Experiment 1A. Sixty-four subjects completed the 2-week version, Experiment 1B. Within each version, an equal number of subjects was assigned to each of the four conditions.

Several students began but did not complete participation in these experiments, and are not included among the 80 described previously. In Experiment 1A, 9 subjects failed to reach acquisition criterion in the hour allotted for original learning; 7 subjects were excluded from Experiment 1B for the same reason. The data of 2 further subjects in Experiment 1B are not included because 1 failed to return for Session 2 and an equipment failure disrupted testing of the other.

An effort was made to have subjects return for relearning precisely 2 or 4 weeks after original learning. However, a few subjects returned 1 or 2 days early or late. The average retention intervals for the two groups remained 2 and 4 weeks. Subjects participated individually in both phases of the experiment.

Materials and apparatus. Each subject learned a 20-item paired associate list during acquisition. The stimulus terms, two-digit numbers of low association value from the Battig and Spera (1962) norms, were from the same association range (from 1.34 to 1.99) used in previous studies (cf. Nelson et al., 1979). The response terms were either concrete nouns or their simple line-drawing picture counterparts, depending upon the condition. These items were selected from the Snodgrass and Vanderwart (1980) norms, with the constraint that 98%–100% of their respondents gave the same verbal label to the pictures. This assured maximal picture–word correspondence.

Two equivalent sets of 20 response items were constructed so that every stimulus term was paired with two potential responses (see Appendix A). Care was taken to avoid any association across pairs of responses. For each subject, 10 items were chosen at random from each response set to comprise the original learning list. One response set was designated as the relearning list for all subjects, with half of the subjects relearning pictures and half relearning words. In this way, the final list was made up of the same stimulus–response pairs for all subjects in all conditions, all manipulations having taken place at original learning. Furthermore, half of the stimuli had identical responses at original learning and relearning (same), and the other half of the stimuli had new responses at relearning unrelated to those at original learning (changed). This procedure avoids confounding the relation between relearning and original learning with the specific items that occur during relearning.

Each paired associate item was prepared as a black-on-white 35-mm slide. Words and numbers were typed in uppercase characters; pictures were photographed from the Snodgrass and Vanderwart (1980) set. Pairs appeared in the form 27—UMBRELLA (or the corresponding picture). A further set of slides of the form 27—? was prepared for use on test trials. Also, a set of slides with rows of random integers was created for use between study and test trials. Finally, a set of four vowel-consonant pairs with appropriate test versions was prepared for the practice sessions. All materials were presented via a Kodak Carousel projector with an on-board timer.

Procedure: Acquisition. Before learning the 20-item paired associate list, the subject had one trial on the four-item vowel-consonant list to familiarize him or her with the timing and procedure. First, the subject studied each pair silently for 8 s. Then, following the last study pair, four consecutive slides filled with random digits appeared, each for 8 s. The subject was required to shadow the digits aloud as rapidly as possible. This interpolated distractor activity was included (a) to prevent the possibility of an item being tested immediately after it was studied (because study and test sequences were randomized independently), and (b) to ensure that items recalled on the subsequent test were stored for long-term retention during study. In the final stage of practice, each pair was tested by presenting only the stimulus vowel and permitting the subject 8 s to say the appropriate response consonant aloud. Subjects were informed that only their first response would be accepted and that responses made after the 8 s had elapsed could not be acknowledged. Subjects were encouraged to guess when unsure.

The same timing, interpolated task, and study–test procedure was used for acquisition of the main experimental list. For picture responses, subjects were instructed to say the appropriate verbal label at recall. If a subject incorrectly identified a picture (e.g., by saying "hand" instead of "glove"), the correct label was provided verbally by the experimenter. This happened very rarely. After each study–test block, the study and test slides for all pairs correctly recalled in that block were removed from the set to minimize overlearning (Battig, 1965). To minimize order effects, study and test slides were randomized anew after each block. Subjects were told that the correct response for each stimulus would be constant despite the varying numbers of pairs from block to block.

A fairly stringent acquisition criterion was used in Experiment 1A. After a subject had been correct on all of the pairs once, the 20-item study set was reassembled and the procedure repeated until the subject recalled all items correctly on a single test of the entire list. For the 64 subjects in Experiment 1B, the criterion was one correct recall of every item, without reassembling the 20-item list afterward. This more lenient criterion necessitated the shorter retention interval in Experiment 1B. Both criteria have been used previously in work on savings for verbal information (cf. Nelson et al., 1979).

Procedure: Retention test and relearning. Subjects had not been told of the retention test phase of the experiment during acquisition, nor were they informed when recruited again by telephone 2 or 4 weeks later. When they returned to the laboratory, they first were refamiliarized with the procedure via the practice list. Following this, they were given a self-paced, forced-response retention test of the originally learned pairs. Self-pacing allowed subjects as much time as they needed to retrieve any items they might remember, even difficult long-latency items. To avoid ambiguity in defining a forgotten item, omissions were not permitted.

Each list was made up of only picture or only word responses in both sessions. Thus, there were four possible relations between the original learning and the relearning lists: word–word, word–picture, picture–word, and picture–picture, where the first element applies to original learning and the second to relearning. This completely between-subjects manipulation was intended to minimize the contribution of differential processing strategies for words and pictures during list learning.

After the retention test for the original pairs, there was a single relearning study–test sequence (cf. Nelson, 1971). Half of the pairs on the relearned list were identical in referent to those on the original list (same), and half had a new, unrelated response paired with an original stimulus (changed), following MacLeod (1976). Study was again at an 8-s rate, followed by 32 s of overt number shadowing. The retention test for the relearning list immediately followed the distractor interval and was self-paced and forced response. Subjects were cautioned to try to remember the pairs from the relearning list, not the original list.

Results

Acquisition and retention. Table 1 displays the data for number of trials to criterion and proportion of items correctly recalled on the delayed retention test. The data are presented separately for Experiments 1A and 1B. Because of the more stringent criterion in Experiment 1A, these 16 subjects took longer to acquire the list, as indexed by the mean number of study trials. They also remembered more responses correctly after the retention interval, presumably because they had learned the list better at the outset.

In Experiment 1A, individual proportions correct on the delayed recall test ranged from .20 to .90 across subjects. In Experiment 1B, the range of proportions correct was from 0 to .65. Only 2 of the 64 subjects in Experiment 1B recalled no items; in the analyses reported next, group means were used to fill their empty cells, but excluding the data for these two subjects did not alter the results. Within each experiment, there were no significant group differences in acquisition or in retention, all $Fs < 1$. The four groups in each experiment were comparable prior to relearning.[2]

Relearning. Table 2 presents the relearning data for both Experiments 1A and 1B. Before detailed examination, a brief description of how the data are organized is worthwhile because the same layout is used in all five experiments. The first major distinction in Table 2 is between items correct and incorrect on the recall test following the retention interval. We would normally call these *remembered* and *forgotten* items, respectively. Because the savings data emphasize the relative nature of the term *forgotten*, and because both recognition and recall tests will be used across experiments, the distinction here will be between *recalled* and *not recalled* items. Use of these test-specific terms should also help the

[2] Postman (1978) also found no long-term retention differences for pictures and words in paired-associate learning. However, he found slower learning with pictures as response terms. Differences in the stimulus terms and other procedural differences between the two experiments may account for this apparent discrepancy.

Table 1
Experiment 1: Mean Number of Trials to Acquisition and Mean Proportion of Responses Correctly Recalled Before Relearning as a Function of Condition

Dependent variable	Condition: Picture–Picture	Picture–Word	Word–Picture	Word–Word	*M*
Experiment 1A: 4 week ($n = 16$)					
Trials to acquisition	8.00	8.50	8.00	8.75	8.31
Correct delayed recall	.61	.48	.51	.45	.51
Experiment 1B: 2 week ($n = 64$)					
Trials to acquisition	4.81	5.06	4.75	5.06	4.92
Correct delayed recall	.23	.28	.28	.24	.26

reader keep in mind which test results are currently being considered.

Intuitively, the recalled and not recalled items are two rather different collections of items. If an item is recalled correctly on the delayed test, that same item certainly should be relearned with very high probability. After all, no new learning need take place. The situation is quite different for an apparently forgotten item. To be remembered after relearning, some new learning must take place. The interesting question in the case of the not recalled item is whether something is left in memory to assist relearning.

The second major distinction in Table 2 derives from the manipulation on the relearning trial. Contrasting same items with changed items, whether in the recalled set or the not recalled set, provides a measure of the extent to which relearning the original item (or a related item) is superior to learning an unrelated item. Thus, the changed condition constitutes a baseline (control) condition against which to measure savings. In a way, a relearning advantage for same over changed items in the recalled set is a precondition for evaluating the crucial not recalled set. If a remembered item cannot be relearned better than a new item, one would hardly expect an unremembered item to be relearned better than a new item. Thus, it is always worthwhile to examine relearning performance on the recalled items first. However, the major focus of savings research is on the not recalled items.

Table 2
Experiment 1: Proportions of Correctly Recalled Words and Simple Pictures Following Relearning, Separately for Recalled and Not Recalled Items

Relation of relearning item to original item	Condition: Picture–Picture	Picture–Word	Word–Picture	Word–Word	*M*
Experiment 1A: 4 week ($n = 16$)					
Recalled					
Same	.84	1.00	.95	1.00	.95
Changed	.38	.23	.37	.13	.28
Not recalled					
Same	.70	.85	.92	.90	.84
Changed	.32	.43	.58	.22	.39
Experiment 1B: 2 week ($n = 64$)					
Recalled					
Same	.79	.85	.97	.96	.89
Changed	.49	.38	.34	.38	.40
Not recalled					
Same	.55	.53	.58	.71	.59
Changed	.34	.34	.30	.31	.32

The following analyses are 2 × 4 mixed analyses of variance with a within-subjects variable of item relation (same vs. changed) and a between-subjects variable of list type (picture–picture, picture–word, word–picture, and word–word).

Analysis of variance of the relearning data for recalled items in Experiment 1A demonstrated that same items were relearned significantly better than changed items, $F(1, 12) = 214.75$, $MS_e = .016$, $p < .001$. Although the four groups did not differ from each other, $F < 1$, the Item Relation × List Type interaction was significant, $F(3, 12) = 3.90$, $p < .05$. Given only 4 subjects per list type, it is probably best to attribute this interaction to the fluctuations apparent over the changed items, which did not form a meaningful pattern.

With its larger sample size, the picture for recalled items was clearer in Experiment 1B. Only the effect of item relation was significant, $F(1, 60) = 92.28$, $MS_e = .085$, $p < .001$. The four groups did not differ, $F < 1$, nor did these two variables interact, $F(3, 60) = 1.97$, $p > .10$. Thus, both experiments showed significantly better relearning of same versus changed responses for recalled items, regardless of the relation between original list type and relearning list type. This demonstrates the representativeness of the set of materials used.

The data of primary concern are those for items not recalled on the retention test prior to relearning. There should be savings for identical words; will savings appear when pictures make up the materials in one or both lists? From both experiments, the answer was *yes*. For Experiment 1A, the effect of item relation was highly significant, $F(1, 12) = 23.01$, $MS_e = .072$, $p < .001$, with performance on same items ($M = .84$) much higher than on changed items ($M = .39$). List type and item relation did not interact, $F < 1$, but there was a tendency toward an overall difference between the groups, $F(3, 12) = 2.63$, $MS_e = .033$, $.05 < p < .10$.

Experiment 1B produced results for not recalled items that were consistent with those of Experiment 1A. The significant effect of item relation, $F(1, 60) = 77.56$, $MS_e = .030$, $p < .001$, again demonstrated strong savings. Performance on the same items ($M = .59$) was considerably better than performance on the changed items ($M = .32$). The nonsignificant effect of list type, $F < 1$, indicated similar performance over the four groups, although this was qualified by a marginally significant interaction, $F(3, 60) = 2.33$, $.05 < p < .10$. There was a tendency toward greater savings when the original item was a word as opposed to a picture, particularly if the relearning item also was a word.[3] Still, all four groups evidenced

[3] If we compare the word–word condition with the mean of the

reliable savings for apparently forgotten items in both versions of Experiment 1.

Discussion

These results, and their interpretation, are quite straightforward. Even when a picture or a word that was learned weeks ago cannot be recalled, information relevant to that item persists in memory. This information can assist the relearning of the identical item or of its counterpart in the other modality. Furthermore, the boost provided by sharing the same referent occurs regardless of modality, although it may be greater for words than for pictures or combinations. Still, savings occurs when the two modalities are identical (both pictures or both words) and when they are different (one a word and the other the corresponding picture).

The savings observed for nominally forgotten items in the crossover conditions (i.e., word–picture and picture–word) corresponds nicely with the cross-language results for English–French bilinguals (MacLeod, 1976). Yet studies by Nelson (1971, Experiment 2) and by Nelson et al. (1979, Experiments 1, 4, 5, and 6) repeatedly have demonstrated the absence of a relearning advantage for synonyms of forgotten responses (despite a marked advantage for synonyms of remembered responses).[4] Although a picture and a word, or words in two languages, that point to the same referent do show savings, two words in the same language thought to point to the same referent do not show savings. This is a perplexing problem, and will be raised again later in discussing the results of Experiment 3. For the moment, it should simply be noted that such results are at odds with the levels of inclusiveness hypothesis (Nelson et al., 1979), which predicts no savings for items at the same level.

Of course, it would be premature to generalize widely from the present experiments, which are the first in the domain of savings for nonverbal material. Two features of Experiments 1A and 1B particularly limit broad inferences. First, all testing for both pictures and words at all stages of the experiment involved recall of verbal labels. It is possible that subjects were recoding pictures into words so that they could meet this verbal recall requirement. Experiments 2 and 3 will address this possibility. Second, the fact that all pictures were simple line drawings chosen specifically so that they had easily and reliably accessible labels also may have encouraged recoding. Whether similar results would obtain for complex, difficult-to-recode pictures remains to be seen. Experiments 2 and 4 will be directed at this question. For the moment, Experiment 1 has demonstrated that it is worth pursuing the study of savings for pictures.

other three conditions, we find, respectively, .68 versus .38 in Experiment 1A and .40 versus .23 in Experiment 1B. Both suggest better retention of words than pictures when savings is the measure, which is usually not the case in recall or recognition. Interestingly, Weldon and Roediger (1987) reported a similar reversal in examining priming in word fragment completion: Pictures produce less priming than words. Of course, some caution must be exercised in interpreting the present Experiment 1A because of its small sample size and the fact that the changed condition is lower rather than the same condition being higher in the word-word case.

Experiment 2

The principal idea motivating Experiment 2 was to bring together relearning/savings and the huge literature on recognition memory for pictures, which typically uses complex, naturalistic scenes as the to-be-remembered material (cf. Loftus, 1982). For optimal generalizability, this first encounter between the two areas should retain as many characteristics of each as possible. Thus, the materials were color photographs of natural scenes, and recognition tests were used throughout this experiment. One new issue was raised by switching to recognition of complex pictures. Because picture recognition is known to be exceedingly good in most situations (e.g., Shepard, 1967; Standing, 1973), it was hard to know what length of retention interval to use. Accordingly, five separate replications of the experiment were carried out with retention intervals of 2, 4, 6, 8, and 10 weeks.

Two further concerns were addressed by this second experiment. First, the fact that the pictures in Experiment 1 were readily labeled simple line drawings permitted verbal recoding, and may undermine conclusions with respect to savings for pictures in general. Second, the fact that subjects recalled verbal labels for pictures in Experiment 1 probably discouraged attention to the visual aspects of the pictures. Use of naturalistic scenes and recognition tests should alleviate these concerns. These changes meet Nelson's requests to broaden the domain of savings research by studying a broader range of materials (Nelson et al., 1979, p. 247) and by using recognition tests during acquisition (Nelson, 1978, p. 466).

The principal intention of Experiment 2 was to determine whether there is savings for apparently forgotten (unrecognized) complex pictures. Until now, it could be argued that only words and verbally recodable drawings have shown savings (and then only using recall tests), so Experiment 2 was seen as an important step. A secondary goal was to determine an appropriate retention interval for future studies of savings for complex pictures.

Method

Subjects. A total of 80 undergraduates at the Scarborough Campus of the University of Toronto participated for course credit. Subjects were assigned to one of the five groups defined by retention interval until there were 16 subjects in each group. Three subjects who failed to return for the second session were replaced. All testing was done individually, and all subjects returned for their second session on the exact day specified by their retention interval.

Materials. Each subject learned a list of 20 number–picture paired associates. The number stimuli were the same as in Experiment 1. The pictures were 3.5 × 5-in. (8.9 × 12.7 cm) color prints of naturalistic scenes photographed from several photography magazines and from the *National Geographic.* From an original set of approximately 150, 34 were chosen for use in the experiment. The main criteria for selection were (a) that no picture was too visually similar to any other picture, and (b) that no picture could be recoded verbally

[4] In an unpublished study using 16 of the best pairs of noun synonyms tabulated by Whitten, Suter, and Frank (1979), I have also failed to observe savings for synonyms, replicating Nelson's findings.

very easily. Of course, both of these are subjective judgments, but two raters agreed on the final set of pictures.

Of the 34 pictures, 4 were reserved for use as examples. Of the remaining 30 pictures, 20 were paired, one each, with the 20 number stimuli. These would make up the relearning list for all subjects. With the remaining 10 pictures and a randomly selected 10 of the stimulus numbers, another set of pairs (Set A) was constructed. A final set (Set B) was made up of the same 10 pictures used in Set A paired with the remaining 10 stimulus numbers not used in Set A. At acquisition, half of the subjects in each retention interval group learned Set A plus the 10 pairs with different numbers from the relearning set. For the other half of the subjects, Set B was combined with the 10 pairs with different numbers from the relearning set. In this fashion, each subject had 10 same and 10 changed pairs at relearning, but the manipulation took place at acquisition.

For study, each pair was assembled on a 5 × 8-in. (12.7 × 20.3 cm) index card with the stimulus number on the left and the response picture on the right. For test, the number stimuli were each prepared on a 3 × 5-in. (8.9 × 12.7 cm) index card and a second set of prints was used. The four practice vowel-consonant pairs were constructed in the same way. For distraction between study and test, six 5 × 8-in. (12.7 × 20.3 cm) cards were prepared with a different three-digit number on each. All materials were encased in plastic sleeves.

Procedure. The acquisition procedure was similar to that in Experiment 1, with three modifications. First, the study time per pair was reduced to 5 s. Second, distraction between study and test phases involved counting aloud backward by threes from a three-digit number as rapidly as possible (cf. Peterson & Peterson, 1959). Third, all tests during acquisition were recognition tests. The subject was shown each stimulus number together with four possible response pictures. One of these was always the correct response, while the three distractors were selected at random from among the other 19 response pictures, so that all were familiar. The four possible responses were presented as a randomized array. The subject chose by pointing and was allowed as long as needed. The acquisition criterion was the stringent once correct on the entire list. No mention was made of Session 2 at the end of Session 1.

At an interval of 2, 4, 6, 8, or 10 weeks after acquisition, each subject was telephoned again and asked to take part in a "second experiment." After the vowel-consonant practice phase, a single self-paced test of the original pairs was administered. This was again a four-alternative forced-choice recognition test. Next, a single relearning study trial was conducted with a 5-s presentation rate. Because of the manipulation at study, 10 of the pairs were identical (same) and 10 were unrelated (changed) to the pairs in the original list. The single test for the relearning set was a recognition test conducted precisely like the preceding tests.

Results

Acquisition and retention. Table 3 presents the data for trials to criterion and proportion correct on the recognition test following the retention interval. These are displayed separately for the five retention-interval groups. Analysis of variance demonstrated that acquisition rate was equivalent for all five groups, $F < 1$, as would be expected. Also as expected, there was a significant decline in delayed recognition over retention interval, $F(4, 75) = 4.33$, $MS_e = .028$, $p < .01$. This gradual loss was reflected in a significant linear trend, $F(1, 75) = 12.86$, $p < .001$. Both results are as anticipated, although the rate of forgetting of number–picture pairs was in itself of some interest. Apparently, it is quite slow.

Proportions correct on the delayed retention test ranged from a low of 0.15 to a high of 1.00 per subject across the five retention intervals. For the 5 subjects with perfect recognition (2 in the 2-week group and 1 each in the 4-week, 6-week, and 10-week groups), group means were inserted in their empty cells for the reported analyses. The alternative of eliminating all data for these 5 subjects did not alter conclusions in any way.

Table 3
Experiment 2: Mean Number of Trials to Acquisition and Mean Proportion of Responses Correctly Recognized Before Relearning as a Function of Retention Interval

Dependent variable	Retention interval (in weeks) 2	4	6	8	10	*M*
Trials to acquisition	4.56	4.88	4.88	4.94	4.38	4.73
Correct delayed recognition	.80	.68	.72	.58	.61	.68

Relearning. Table 4 presents the relearning data separately for items remembered and forgotten on the delayed recognition test. Each of the retention intervals can reasonably be thought of as an independent replication of the experiment because all displayed adequate forgetting for the analysis of savings.

A 2 × 5 analysis of variance was carried out on the data for recognized items at the top of Table 4. The factors were item relation (same vs. changed) and retention interval. Neither the effect of retention interval nor its interaction with item relation was significant, both $Fs < 1$. However, the highly significant effect of item relation, $F(1, 75) = 14.25$, $MS_e = .012$, $p < .001$, demonstrated a reliable relearning advantage for identical pictures remembered over a long retention interval. The relatively small average size of the effect (7%) may have been due to performance on the same pairs approaching the measurement ceiling. Still, the precondition for examining savings for forgotten items was met.

In the data for items not recognized after the retention interval, a very different pattern was observed. There were no significant differences. As earlier, the effect of retention interval and its interaction with item relation were both nonsignificant, both $Fs < 1.28$. The unexpected result was that the

Table 4
Experiment 2: Proportions of Correctly Recognized Complex Pictures Following Relearning Shown as a Function of Retention Interval, Separately for Recognized and Not Recognized Pictures

Relation of relearning item to original item	Retention interval (in weeks) 2	4	6	8	10	*M*
Recognized						
Same	.96	.95	.91	.91	1.00	.95
Changed	.91	.84	.87	.86	.90	.88
Not recognized						
Same	.91	.94	.83	.82	.79	.86
Changed	.87	.83	.78	.79	.90	.83

effect of item relation was also nonsignificant, $F < 1$, indicating no savings for not recognized items. This is the first savings experiment where an identical response has failed to boost relearning. Is there something wrong with the experiment, or is this absence of savings real?

Discussion

Consider first the possibility that something has gone wrong with the experiment. Certainly, there is greater variance in the not recognized data than in the recognized data. This is confirmed by the differences in MS_es (.012 for recognized vs. .041 for not recognized). But does this greater variance obscure a small effect? Another concern is that performance for the not recognized items is quite high, although it is below that of the recognized items where a significant same–changed difference was observed. The only way to deal with these possibilities is empirically, which is the major goal of Experiment 4.

Consider now the second possibility raised earlier—that there really is no savings for forgotten items here. Why might this be? Two quite different reasons come to mind. First, it may be that only words or verbally codable items can show savings. If a complex picture is recognizable, then that same picture can contact its representation in memory and produce a relearning advantage over a novel picture. However, if the picture is not recognizable, sufficient information may have been forgotten to prevent that contact during relearning. Hence, there will be no relearning advantage for identical (but unrecognized) pictures. Of course, the task then would be to explain why this is more of a problem with pictures than it is with words.

Assuming that the absence of savings is real, the second possible reason relies on the type of test, not the type of material. In moving to complex pictures, a simultaneous move to recognition testing was made. Could it be that savings only appears on a recall test? The only prior uses of a recognition test of savings were by Nelson (1978, Experiment 3) and by Groninger and Groninger (1980). Interestingly, Nelson's recognition test did not detect a difference between the same and changed conditions; Groninger and Groninger's did. This inconsistency must be straightened out. If savings cannot be shown using a recognition test of relearning, this would raise concerns about whether the phenomenon derives solely from trace strengthening. A trace made stronger by relearning should make itself evident on either type of test.

Experiments 3, 4, and 5 pursue the questions raised here. In all three, both recall and recognition tests will be used to evaluate relearning. The material varies from line drawings to complex pictures to words, coming full circle. If there are conditions under which there is no savings, these experiments will permit determination of whether the type of material, the type of test, or some combination of the two is the critical factor.

Experiment 3

Experiment 1 used recall tests to duplicate Nelson et al.'s (1979) procedure with words and simple line drawings. Experiment 2 used recognition tests to follow standard practice in studying memory for complex pictures. There was savings for the simple pictures in Experiment 1 but not for the complex pictures in Experiment 2. To begin to break this confounding of material and test type, Experiment 3 returned to simple pictures as the material and evaluated relearning by both types of test. If verbal codability is the key factor, then simple pictures should permit recoding into verbal form and savings should be observed under both tests. If type of test is the key, and recognition tests cannot reveal savings, then even simple pictures should fail to show a relearning advantage when measured by recognition, but not when measured by recall.

Experiment 3 was designed to be very similar to Experiment 2 with only one alteration other than the material and test format changes. In the new experiment, subjects could relearn one of three pictures: the identical picture, another picture having the same referent, or an unrelated control picture. For example, having learned a picture of a desk telephone originally, the subject could relearn either that same desk telephone, a different-looking wall telephone, or a lion. Taken at face value, the levels-of-inclusiveness hypothesis predicts savings only for the identical picture and not for another picture at the same level, despite sharing a referent. More fundamentally, the recall-recognition contrast will tell us more about the trace-strengthening hypothesis, which does not predict a differential effect of relearning test type.

Method

Subjects. A total of 40 subjects from the same pool as previous experiments participated either for course credit ($n = 10$) or for a $4 payment ($n = 30$). Of the 40, 1 failed to return for the second session, a second wrote down the list and studied it before returning, and there were procedural problems for a third. Four others failed to complete acquisition within the hour allotted, leaving complete data on 33 subjects.

Not all subjects could return precisely 6 weeks after acquisition, so a 2-day grace period was permitted in either direction. Nine subjects returned late ($M = 1.67$ days) and 11 returned early ($M = 1.64$ days) for an average retention interval of almost exactly 6 weeks. Subjects participated individually in both sessions.

Materials. Each subject learned a 24-item paired associate list with numbers as stimuli and line drawings as responses. The numbers were those from the previous experiments plus four more from the same pool. The simple line drawings were selected from the "synonymous" pictures used by Tversky (1979); labels for these pictures are contained in Appendix B. Although the labels in the first two columns are identical, the pictures were in fact physically different realizations of the same referent. Any apparent visual or semantic relation between the same-referent and control pictures associated with each stimulus number was avoided. Each number–picture pair was prepared on a 3 × 5-in. (8.9 × 12.7 cm) index card. For the recognition and recall tests, a set of number-only and picture-only cards was constructed.

Procedure: Acquisition. Following one study–test sequence with the practice list, subjects had a preliminary trial where they named each of the 24 pictures aloud once to insure correct identification. In the few cases of incorrect identification, the subject was corrected. Then acquisition began. Study used a 5-s rate per pair, with the pacing guided by an audible click from a tape recording. The 30 s of distraction between study and test was number shadowing, as in Experiment 1.

Tests were four-alternative forced-choice recognition tests constructed as in Experiment 2. All testing was subject paced, with the subject aware that only the first response for a particular stimulus could be accepted. A response was required for every stimulus, even if the subject eventually had to guess. After each test phase, the study and test cards for all correctly recognized pairs were removed from their respective decks to minimize overlearning. Study and test decks were shuffled independently after every trial. Acquisition criterion was two correct recognitions of each pair.

As in Experiments 1 and 2, item manipulation occurred at acquisition so that the relearning list was constant over subjects. A subject's acquisition list consisted of 24 pairs, with 8 eventually falling into each of the three conditions: identical, same referent, and unrelated (control). In terms of Appendix B, 8 responses were selected at random from each column so that every stimulus had only one response. Each subject had a unique acquisition list. At relearning, 16 subjects learned Picture Set A and 17 learned Picture Set B.

Procedure: Retention test and relearning. No mention of a second session had been made during the first. Six weeks after acquisition, the subjects were telephoned and asked to return for "another experiment." Except for changes already noted, the procedure in the second session was identical to that of Experiment 2 until the end of the relearning study trial. At that point, an unexpected recall test was given using the procedure in Experiment 1. The subject was allowed as much time as needed for each response, with all pairs tested. Following the recall test, a final test of recognition for the relearned items was administered. The same format was used as in all previous recognition tests. All subjects had the recall test prior to the recognition test to avoid the well-established effect of a prior recognition test on a subsequent recall test (e.g., Postman, Jenkins, & Postman, 1948). The reverse order—recall prior to recognition—is less problematic (e.g., Hanawalt & Tarr, 1961; Watkins & Todres, 1978), and is recommended by Brown (1976, p. 34) in his overview.

Results

Mean number of trials to attain the acquisition criterion was 4.54. This is similar to the values observed in Experiment 2. Mean proportion of responses correctly recognized after 6 weeks was .54, making the recognized and not recognized pools of items roughly comparable in size. The range of proportions correct on the delayed retention test was from .17 to .79 per subject.

Table 5 displays the mean proportions of responses correctly recalled and recognized following relearning. This is done separately for items recognized and not recognized on the delayed retention test. Consider first the recognized data. There was a clear relearning advantage here whether measured by recall, $F(2, 64) = 34.26$, $MS_e = .049$, $p < .001$, or by recognition, $F(2, 64) = 4.55$, $MS_e = .034$, $p < .05$. Same-referent pictures appeared to produce virtually the same relearning advantage as identical pictures. These results for the recognized items demonstrated that the materials in this experiment were appropriate.

Now consider the data for not recognized items. As measured by recall, there was marginally significant savings, $F(2, 64) = 2.71$, $MS_e = .086$, $.05 < p < .10$. Both the identical and the same-referent pictures were relearned better than the control pictures by about 15%. The superior relearning of identical line drawings as measured by recall replicated the picture–picture condition of Experiment 1, although the effect was somewhat smaller here. The new result was that nonidentical pictures that point to the same referent in memory also can boost relearning, apparently to the same extent as identical pictures. An intriguing aspect of this finding is that Tversky (1979) considered these same-referent pictures to be "synonymous," yet Nelson (1971, Experiment 2) and Nelson et al. (1979, Experiments 1, 4, 5, and 6) have demonstrated that verbal synonyms do not show savings. Perhaps it is better to reserve the designation "synonym" for the verbal domain.

The crucial result of Experiment 3 concerned the recognition test data for items relearned after being forgotten during the retention interval. The 8% and 2% advantages for identical and same-referent pictures did not differ reliably from the control condition, $F(2, 64) = 1.12$, $MS_e = .049$, $p > .30$. That is, there was no relearning advantage evident for not recognized simple pictures when relearning was indexed by a recognition test. This replicated the finding for complex, naturalistic scene pictures in Experiment 2.

Table 5
Experiment 3: Proportions of Correctly Recalled and Recognized Simple Pictures Following Relearning as a Function of Condition, Separately for Recognized and Not Recognized Pictures

Type of test after relearning	Condition		
	Identical	Same-referent	Control
Recognized			
Recall test	.66	.61	.24
Recognition test	.85	.83	.72
Not recognized			
Recall test	.44	.43	.29
Recognition test	.83	.77	.75

Discussion

Three main findings emerged from this experiment. First, when measured by recall, there was savings for simple pictures. This replicated Experiment 1 and generalizes the result to cases where acquisition is by recognition. Second, different pictures that share the same referent also can show a relearning advantage, at least when measured by recall. And third, when measured by recognition, neither identical nor same-referent pictures evidenced a significant relearning advantage over unrelated control pictures.

The first finding was expected, of course, because it was basically a replication. However, it is useful to know that the nature of the test during acquisition has little impact on subsequent relearning (see also Groninger & Groninger, 1980, and Nelson, 1978, p. 466). The second finding is more novel. As mentioned, synonymous words do not show savings, but same-referent pictures do. Why? Two related possibilities seem feasible. One is that same-referent pictures give rise to the same word, which synonyms obviously do not. The information left in an unrecalled (or unrecognized) memory trace may include the nuances of meaning that differentiate synonyms. Of course, this argument assumes that subjects are recoding simple pictures into their verbal counterparts. Consistent with this assumption is the observation that different pictures of the same referent produce equivalent savings when measured by recall.

The second possibility is that similar pictures give rise to the same *concept* in memory, unlike synonymous words. This receives some support from the work on savings for translation equivalents in bilinguals (MacLeod, 1976, Experiment 1). Here, the language of the relearned item made little difference; as long as the meaning (or referent) was preserved, savings was comparable within and between languages. Like the word-based hypothesis above, this concept-based hypothesis relies on the idea that pairs of verbal synonyms have different, though overlapping, referential fields.

At present, the most central of the three new findings summarized earlier is the recall-recognition contrast. In two attempts, a post-relearning test of recognition has failed to reveal significant savings. This is true despite the fact that use of the same stimuli in the pairs should, if it has any effect [cf. Footnote 2], cause interference with learning changed items and thereby enhance apparent savings. The implication is that it is the type of test, not the type of material, that accounts for the absence of savings for pictures in Experiments 2 and 3. This is emphasized by the fact that a recall test did reveal savings for simple pictures while a recognition test did not in Experiment 3. To put this hypothesis to a still more stringent test, Experiments 4 and 5 will use both recall and recognition to measure savings for complex pictures and words.

Experiment 4

In the last two experiments, neither simple line-drawing pictures (Experiment 3) nor complex naturalistic pictures (Experiment 2) have shown significant savings on a recognition test following relearning. However, measured by recall, there are two demonstrations of savings for simple pictures (Experiments 1 and 3). To complete the story, what is required is a *recall* test of savings for complex pictures. Recognition tests may not reveal savings in general, but will any test demonstrate savings for complex pictures?

In Experiment 4, a new and larger set of complex pictures was assembled to increase the generalizability of Experiment 2. On the basis of that experiment, a 7-week retention interval was chosen to assure sufficient forgetting. To provide the strongest test of the recall-recognition difference in a single experiment, both types of test were administered.

Method

Subjects. A total of 34 undergraduates from the usual pool volunteered to participate. The data of 1 subject were discarded because he did not learn the list in the allotted hour. Subjects were tested individually in both sessions. Although all subjects were to have returned exactly 7 weeks after acquisition for retention tests and relearning, 4 people returned 1 day late, 4 returned 1 day early, and 1 returned 2 days early.

Materials. Each subject learned a 28-item paired associate list. The stimulus numbers were the 24 from prior studies plus 4 more from the same range. The response pictures were a subset of those used by Tulving (1981), including photographs of rural and urban scenes, people, animals, and objects. An effort was made to choose quite differentiable pictures forming a fairly distinct set. Each number was paired with two target pictures, the first a member of List A and the second a member of List B. In all other respects, materials were prepared as in Experiment 2.

Procedure. After one practice trial with the vowel-consonant pairs, list acquisition began. Rate of presentation, distraction, and recognition testing all were the same as in Experiment 2. The acquisition criterion was two correct recognitions of each pair, as in Experiment 3. For each subject, 14 responses were selected from List A and 14 from List B so that there was one response picture for each stimulus number. A different random set was chosen for each subject. All subjects relearned the entire set of List B responses.

Subjects returned for a "second experiment" 7 weeks after original learning. After one study–test trial on the practice list, a self-paced delayed recognition test for the original responses was given. Then the relearning study trial was conducted using the 28 List B response pictures for all subjects. The self-paced recall test followed, with subjects asked to describe in a few words the picture associated during relearning with a given number. Scoring these protocols was quite straightforward, as subjects' descriptions of pictures were very clear. At no time prior to this recall test had subjects ever been asked to describe the pictures. Finally, following the recall test, a standard recognition test was administered for the relearned items.

Results

Mean number of trials to acquisition was 5.1. This was slightly longer than the previous two experiments, probably due to the slightly longer list. The mean proportion of response pictures correctly recognized on the 7-week delayed recognition test was .46. The longer list and more lenient acquisition criterion probably account for the lower accuracy here than in Experiment 2. Proportions correct on the delayed recognition test ranged from .25 to .79 for individual subjects, and there were no empty cells in the data matrix.

Table 6 presents the relearning data separately for the recall and recognition tests of the recognized and not recognized items. Consider first the recognized items. As in all previous experiments, there was a significant relearning advantage for same items over changed items when those items were retained over the retention interval. This was true for recall, $F(1, 32) = 58.81$, $MS_e = .048$, $p < .001$, and for recognition, $F(1, 32) = 17.45$, $MS_e = .024$, $p < .001$. If a complex, naturalistic scene picture can be recognized correctly after a long retention interval, that picture can be relearned better than a new, unrelated picture. This advantage was evident

Table 6
Experiment 4: Proportions of Correctly Recalled and Recognized Complex Pictures Following Relearning, Separately for Recognized and Not Recognized Items

Type of test after relearning	Item relation	
	Same	Changed
Recognized		
Recall test	.61	.19
Recognition test	.85	.69
Not recognized		
Recall test	.33	.18
Recognition test	.72	.65

regardless of whether a recall test or a recognition test was used as the post-relearning measure. Again, the set of materials can be deemed representative.

As always, the not recognized data were of primary interest here. As measured by recall, there was significant savings for complex pictures, $F(1, 32) = 13.25$, $MS_e = .026$, $p < .001$. Identical pictures were relearned about 15% better than unrelated pictures according to the recall test. Thus, complex pictures can benefit from relearning, even when they have just failed to be recognized. This is the first demonstration of a savings advantage for complex, naturalistic pictures.

Unfortunately, the post-relearning recognition test results confuse the story. There was also some suggestion of a savings advantage as measured by recognition, although it was only marginally significant, $F(1, 32) = 3.64$, $MS_e = .021$, $.05 < p < .10$. In the face of this almost reliable 7% difference, it would be premature to conclude that savings can never be demonstrated via a recognition test (see also Groninger & Groninger, 1980). Experiment 5 was aimed directly at the recall-recognition contrast in an effort to resolve the issue.

Discussion

The principal new finding of Experiment 4 was that a relearning advantage could be demonstrated for complex pictures, at least if a descriptive recall test was used after relearning. The recall data clarify the outcome of Experiment 2 by showing that identical relearning responses always result in savings, independent of the nature of the material, so long as the measure of relearning is recall. By inference, the failure to find savings for identical complex pictures in Experiment 2 is more probably due to the use of only a recognition test as a measure.

Experiment 5

The pattern emerging thus far is quite consistent, with the exception of the recognition data in the previous experiment. If a recall test is used to evaluate relearning, a reliable amount of savings will be observed whether the items are identical words, simple pictures, or complex pictures. This is true in all of the previous word studies, as well as in the three new picture experiments (Experiments 1, 3, and 4). On the other hand, if a recognition test is used to evaluate savings, three new studies (Experiments 2, 3, and 4) have failed to demonstrate reliable savings for simple and complex pictures.

Is it the case that a recognition test following relearning simply does not work as an index of savings? This would be very much at odds with the predominant trace strengthening view of relearning (Ebbinghaus, 1885; Nelson et al., 1979), which does not predict test differences. Perhaps the strongest conceivable test of this hypothesis is to complete the materials by type of test matrix. To date, only two experiments (Groninger & Groninger, 1980; Nelson, 1978, Experiment 3) have attempted to evaluate relearning of words via a recognition test. Although Groninger and Groninger obtained reliable savings as measured by a recognition test, Nelson did not. This issue must be resolved. This is the goal of Experiment 5, which uses words as the material and includes both recall and recognition tests of relearning.

Method

Subjects. The subjects were 25 undergraduates from the same pool as the preceding experiments. All received credit toward their introductory course grades for participating. The data of two subjects were discarded: One did not return for the second session, and another did not learn the list in the allotted hour. Of the 23 remaining subjects, 1 was tested a day early, 1 a day late, and 1 two days late in the second session. All subjects were tested individually in both sessions.

Materials. The original learning list was made up of 28 number–noun paired associates. The stimulus numbers were identical to those in Experiment 4 and the nouns were chosen from the Snodgrass and Vanderwart (1980) norms. As Appendix C shows, many of the nouns used in Experiment 1 were used here also, to provide greater continuity. Two sets of 28 responses were constructed in such a way that the two nouns connected to a given number appeared to be unassociated.

Procedure. Acquisition was conducted exactly as in Experiment 4 except that half of the subjects relearned List A and half relearned List B. By counterbalancing over pairs of subjects, all responses occurred equally often in all three conditions: original learning only, relearning only, and both. All tests during acquisition were recognition tests. The acquisition criterion was two correct recognitions of each of the 28 responses, as in Experiments 3 and 4.

Subjects returned after 6 weeks for an unexpected retention test and relearning. This, too, was carried out precisely as in Experiment 4. The only other change made in Experiment 5 was that the entire procedure was controlled by an Apple II+ microcomputer, which required that subjects type in their recall and recognition responses at the keyboard.

Results

The mean number of trials to the acquisition criterion was 6.3. This was longer than in the previous experiment, suggesting that words were a little harder to learn than pictures. The range of .21 to .75 ($M = .45$) in proportion of response words correctly recognized per subject after 6 weeks was, however, quite comparable to that for pictures in Experiment 4.

Table 7 displays the relearning results. For recognized items, the same–changed contrast was significant for recall, $F(1, 22) = 49.14$, $MS_e = .031$, $p < .001$, and for recognition, $F(1, 22)$

Table 7
Experiment 5: Proportions of Correctly Recalled and Recognized Words Following Relearning, Separately for Recognized and Not Recognized Words

Type of test after relearning	Item relation	
	Same	Changed
Recognized		
Recall test	.48	.11
Recognition test	.81	.66
Not recognized		
Recall test	.32	.12
Recognition test	.61	.62

$= 5.82$, $MS_e = .042$, $p < .05$, as in all prior experiments. Thus, the materials are suitable for investigating savings.

For not recognized items, the recall results replicated prior studies using words as responses. There was significant savings as measured by recall, $F(1, 22) = 22.42$, $MS_e = .018$, $p < .001$. The critical data concern recognition, and they were very clear. As measured by recognition, there was no evidence of savings even for identical words, $F < 1$.

Discussion

The major finding of Experiment 5 is the absence of savings for words when measured by recognition. At one level, this is not novel; recognition has failed to reveal reliable savings in all three previous attempts here. What is novel is that this failure occurred with words, not pictures, as the response materials. Although Groninger and Groninger (1980) did report reliable savings for words as measured by recognition, their single experiment now is the only exception to the general finding of no significant savings on recognition tests. Perhaps their use of a batch recognition test was responsible, in that such a test where all possible responses are present simultaneously may allow an overall sorting strategy not possible in the present experiments.

It is noteworthy that the only other word study using recognition (Nelson, 1978, Experiment 3) produced a result analogous to the present experiment. (Because his interest was in comparing the sensitivity of recognition and relearning as general measures of memory [where relearning was measured by a recall test], he did not examine the recall-recognition contrast which is the focus here.) On the basis of all of the savings literature, then, it now seems reasonable to offer the following generalization: *Savings for apparently forgotten information is evidenced by recall tests but not by recognition tests, independent of the form of the information.* This generalization holds regardless of whether the "forgotten" information is defined by recall or by recognition tests prior to relearning. It is a surprising statement given the traditional trace strengthening account of relearning.

General Discussion

From the five experiments reported in this article, four new findings emerged: (a) Reliable savings occurs for simple and complex pictures, but this is only evident when a recall test is used to evaluate relearning; (b) the relearning advantage in recall extends to nonidentical pictures that share a common referent; (c) the relearning advantage in recall occurs regardless of whether the modalities (picture vs. word) of the original learning items and the relearning items are the same or different; and (d) there is no reliable evidence of savings for words, simple pictures, or complex pictures when recognition is the measure used to evaluate relearning, whether "forgotten" items are defined by a recall or a recognition test. How should these findings be integrated into our understanding of savings in particular and memory in general? I will begin by considering their relation to the domain of savings in long-term memory, and then work outward to their broader implications for memory.

The Trace-Strengthening Hypothesis

As set out in the introduction, the traditional view of relearning and savings places emphasis on the memory trace and its relative strength in storage (Ebbinghaus, 1885; Nelson et al., 1979). A nonrecalled or nonrecognized item presumably has low (but nonzero) trace strength. For related or identical items, the strength remaining from original learning is assumed to combine with the strength due to relearning. The resulting total trace strength may exceed the performance threshold, in which case the relearned item is remembered. Of course, the trace strength of an unrelated item cannot assist relearning. This view of savings certainly is the most intuitive one; indeed, no alternative position has been suggested since the time of Ebbinghaus.

Although not expressly designed to test the trace-strengthening account of savings, the present study nevertheless calls this account into question. If the concatenation of trace strengths is greater for an item identical at original learning and relearning than it is for two unrelated items, this ought to influence performance on both recall and recognition (as long as ceiling and floor problems are avoided). This is particularly the case if recognition is viewed as a more sensitive index of memory than is recall (e.g., Nelson, 1978; Postman et al., 1948). Yet in four experiments, the savings evident on the recall test of relearning was not evident on the recognition test for either words or pictures. The question, then, is *why* savings is test specific.

The alternative I wish to suggest is that relearning has its major impact on *retrieval*, not on storage. Although retrieval unquestionably is important in both recognition and recall, its role is different in the two types of retention test, a point that has been made often (Gillund & Shiffrin, 1984; Humphreys, 1978; Kintsch, 1977; Mandler, 1979; Tulving, 1976). In arguing for a two-process "generate–recognize" view of recall, Kintsch (1977, p. 265) said, "Retrieval schemes do not play the same role in recognition as in recall." Arguing against the two-process view, Tulving (1976, p. 67) nevertheless also emphasized this retrieval-based distinction: "Recall and recognition are different inasmuch as the retrieval information that is present at the time of recall is different from that at the time of recognition." Thus, the argument to be made here is not original; its only novelty is in its application to tests of *relearning.*

It is easiest to make this argument in the context of a "generate–recognize" model (e.g., Anderson & Bower, 1972; Kintsch, 1977), although such a model is not essential to the account. Assume that, for a same pair, a relearning trial assists in relocating the "forgotten" trace, thereby making it easier to find on a subsequent test of relearning. This cannot happen for a changed pair. Thus, the association for the same pair will be easier to generate after relearning. Successful recall relies heavily on successful search, but successful recognition is less affected by ease of retrieval. Consequently, a post-relearning test of recall will be more influenced than will a corresponding test of recognition.

This alternative to the trace-strengthening view of savings may be called the *retrieval facilitation hypothesis* of relearning. It is not a wholly unprecedented explanatory idea; in fact, Groninger and Groninger (1980) seem to have had a similar idea. The arguments made here can be related to those made by Wickelgren (1979) in distinguishing vertical from horizontal associations. The parallel would be that relearning affects vertical associations more than horizontal ones, and that the former are particularly important in recall. As another example, Schacter (1985) distinguished between nested and unitized structures in accounting for intact priming in amnesia. For a nested structure to be retrieved, "global context" information must be available; this is where explicit remembering fails in amnesics. Unitized structures do not depend on contextual information and are thought to function normally in amnesics. Relearning, I suggest, influences primarily nested structures. Note that these arguments apply quite well to the paired-associate situation, where Humphreys (1978) argued that retrieval processes may not depend on context in recognition.

Nor is the empirical observation of a difference in recall but not in recognition unprecedented. The savings paradigm has much in common with the transfer paradigms of interference theory, and parallel findings exist in that literature. Some effects found in recall (cf. McGovern, 1964) do not appear in recognition (e.g., Bower & Bostrum, 1968; Postman & Stark, 1969). Following Bahrick's (1979) logic, it would appear that the relearning task imposes divergent processes on recall and recognition performance. The suggestion here is that the primary difference is in retrieval. As Tulving (1983, p. 319) has said, "increase in the relevant retrieval information . . . renders the recall of the 'forgotten' even possible again."

By its nature, the savings paradigm may provide the subject with helpful retrieval cues. All manipulations of items between original learning and relearning are applied to both remembered and forgotten items. Typically, there are only three possible item relations: identical, similar in some well specified way, and unrelated. Thus, it is quite possible for the subject to derive the rules for item relations from the remembered set during relearning. Subsequently, these overall, list-wide rules can provide the basis for a retrieval rule. Miller (1958) has demonstrated that existence of such rules improves recall performance. Although controversial, Kintsch (1968) has shown in contrast that the improvement visible in recall disappears in recognition. These results also are very analogous to the savings results.

The remaining question is why the retrieval facilitating effect of relearning applies to some item relations and not to others. As a quick review, Table 8 presents those relations that show savings at the top and those that do not show savings at the bottom. Probably the most direct test of the retrieval hypothesis would be that no relation between items should show savings under the standard forced-choice recognition test of relearning. This has already been shown to be the case for the first four relations at the top of Table 8; it should also be true for the four that remain above the line. On the other hand, if the retrieval demands of a recognition test could be increased, that recognition test might begin to reveal savings for relations above the line in Table 8.

Other less direct tests of the retrieval hypothesis also are possible. For instance, a variety of original learning to relearning item relations could be included in a single list. Then, following relearning, cues of the sort "synonym," "category name," and so on could be provided to assist subjects in recovering items. Under the retrieval hypothesis, these cues might be expected to accentuate savings. It might even be possible to demonstrate savings for the relations at the bottom of Table 8 under these conditions. More speculatively, response latencies may be worth exploring in the savings paradigm because they may provide a particularly sensitive index of retrieval manipulations (cf. MacLeod & Nelson, 1984).

Table 8
Item Relations at Original Learning and Relearning for Which Reliable Savings Occurs and Does Not Occur on a Recall Test of Relearning

Relation	Original learning response	Relearning response
	Reliable savings	
Identical word	butterfly	butterfly
Identical picture (simple or complex)	BUTTERFLY	BUTTERFLY
Different modality (word vs. picture)	butterfly	BUTTERFLY
Same-referent for two pictures	BUTTERFLY (1)	BUTTERFLY (2)
Different language	horse	cheval
Superordinate	vehicle	car
Subordinate	Buick	car
Homophone	tax	tacks
	No reliable savings	
Synonym	house	home
Associate	chair	table
Antonym	hates	loves

Note. Items in uppercase represent pictures; items in lowercase represent words.

Before turning to consideration of the levels of inclusiveness hypothesis, I should note a problem with this retrieval account. When dealing with paired associates, both recall and recognition tests rely on access to the association. Recovery of this association is essential to successful performance on either test. Yet "recovery" is a retrieval operation, so it is difficult to see why the retrieval boost due to relearning would affect recall more than recognition. Thus, the retrieval hypothesis is not pinned down well, and I offer it largely as an alternative to, not a replacement for, the trace-strengthening view. Basically, it would appear that a better general theory of the relation between recall and recognition is needed to explain a difference of the sort observed here. Such a theory would presumably clarify the role that retrieval processes play in both types of retention test.

The Levels-of-Inclusiveness Hypothesis

The present experiments also address the levels-of-inclusiveness proposition offered by Nelson et al. (1979). This proposition asserts that, for items that are not semantically identical, the information saved in the memory trace of a nominally forgotten item incorporates hierarchically related components of meaning but not same-level components of meaning. This accommodates both the observation of reliable savings for superordinates and subordinates and the finding of no reliable savings for antonyms, associates, or synonyms. In 1979, its only apparent shortcoming as a summary statement of the available data was the reliable savings for translation equivalents in bilinguals (MacLeod, 1976, Experiment 1), but this finding can be handled by assuming a common semantic referent as mentioned in the introduction.

As Nelson et al. (1979, p. 247) stated, "more semantic relationships at the same versus different level as the originally learned items should be investigated to test the same-level hypothesis." This was partly the intention of Experiments 1 and 3. In Experiment 1, the same-level items were simple pictures and their corresponding words. In Experiment 3, the same-level items were two different simple pictures of the same referent. Setting aside the data from the recognition test of relearning, there was evidence on the recall test for savings in both experiments. Yet a word and a picture or two pictures for the same referent presumably should have the same level of inclusiveness, just as noncognate translation equivalents should for bilinguals. How are we to reconcile these three findings with the Nelson et al. (1979) proposition?

There seem to be two ways to deal with this situation. The first way is to conclude that the proposition is wrong, in that it is not sufficient in scope to cover at least certain same-level relations. The second way is to add an assumption that permits the proposition to cover all of the available data. One possible modification would be to increase the size of the set of "identical" elements by defining identity in terms of a shared hypothetical concept (such as the idea of a chair) rather than in terms of the physical stimulus (the word *chair* or a picture of a particular chair). This was the basic idea I offered to account for bilingual savings (MacLeod, 1976). A second possible modification would be to suggest that subjects simply think of the appropriate English word whenever they see a corresponding simple picture or a French translation equivalent. This strategy would have the effect of producing identical internal symbols, and all of the data would fit the proposition without recourse to a hypothetical concept.

There are, however, two problems with these amendments of the proposition. First, Nelson et al. (1979, Experiment 6) have shown that there is no savings for synonyms even when subjects are forced to generate synonyms during acquisition. Yet synonyms should point to the same word or concept. Thus, the absence of savings for synonyms returns as a problem for these revisions. The second problem concerns the testability of the proposition so revised. Given the set of same-level relations already examined (see Table 8), it is difficult to think of other semantic relations—whether same-level or different-level—that remain to be tested. It is worth noting, though, that changes in surface form—from visual to auditory, from one typography to another, and so on—should not matter.

With the preceding analysis in mind, the levels-of-inclusiveness proposition may not be the optimal account for the pattern of savings. An alternative account can be derived from the retrieval facilitation ideas presented earlier. In the paired-associate task, the subject must learn to produce the exact words desired by the experimenter as responses. Part of the difficulty, then, is learning to exclude related words from the response set. The argument is that such excluded responses will not show savings because they do not form part of the encoded trace and therefore cannot assist in retrieval of the trace. This idea may be called the *response exclusion hypothesis*.

What evidence is there in favor of this hypothesis? The very words that would be most likely to intrude—synonyms and associates, in particular—did not show savings, and subjects were much slower to acquire the list when forced to generate synonyms on every acquisition trial (Nelson et al., 1979, Experiment 6). For the bilingual case, where reliable savings was observed, an additional assumption may be necessary. Because subjects remember input language extraordinarily well (MacLeod, 1976, Experiment 1), they may perceive no danger of a translation error, so that translation equivalents need not be excluded. Similar reasoning can be applied to all of the remaining relations in Table 8.

Although admittedly very ad hoc, this exclusion hypothesis is not totally implausible, and can be tested quite readily. Consider superordinates which do show savings on a recall test of relearning. This may be the case because none of the responses in the set are superordinates (cf. Nelson et al., 1979, Experiment 2). As the argument goes, if exclusion of superordinates were made necessary during acquisition, savings for them should disappear during relearning. If half of the responses on an acquisition list were basic level and half were superordinates, the subject could not discount superordinate responses as easily and would have to exclude them. Or, if several responses on the list belonged to the same category, superordinate information also might have to be excluded during encoding. Both of these tactics should reduce or eliminate savings for superordinates, and could be tried with any other relation.

The reader will recognize that the exclusion hypothesis really is not a major departure from Nelson's levels-of-inclusiveness hypothesis. As well, it should be repeated that the exclusion hypothesis at present is completely *ad hoc* and requires additional assumptions to cover all of Table 8. The reasons for introducing this new hypothesis are that it makes testable predictions and that it is in the spirit of a retrieval view of savings.

Broader Implications for the Study of Memory

Nelson et al. (1979, pp. 244–247) already have provided a lucid account of the place of relearning/savings research in the study of memory. They detailed the relation of savings work to studies of false alarms in recognition, of intrusions in recall, of positive transfer in learning successive lists, of retrieval cue effects in recall, and of a variety of short-term memory tasks. To this list can also be added the feeling of knowing and metamemory (Nelson et al., 1984). Rather than reiterate their discussion, I will confine my comments in this section to the implications of savings research for four specific issues: memory for pictures, recognition versus recall, the distinction between implicit and explicit tests of memory, and the nature of representation in memory.

The original aim of this article related to the first issue—bringing together the savings methodology with the study of memory for pictures. At least when recall tests of relearning are used, the present experiments demonstrate that this may well be a fruitful combination. Relearning provides another tool for the study of picture memory, one that may be ideal for investigating information remaining intact over periods of months (e.g., Mandler & Ritchey, 1977). Relations concerned with shape, orientation, and other visual properties of pictures (and, indeed, words) remain to be explored in the savings paradigm and may provide evidence converging on the traditional recognition data.

The second issue is the relation between recall and recognition as measures of savings. From the days of Ebbinghaus until very recently, the measurement of savings has always involved producing the previously learned material, in the manner of a recall test. The present study attempted to extend the study of savings to recognition tests of relearning. The naive expectation might have been that recognition tests would respond similarly to recall tests, but this appears not to be the case. Recognition is not simply a more sensitive memory test than recall; rather, the two types of tests generally seem to rely on different aspects of retrieval. Future studies of savings should continue to examine the role of the test following relearning, perhaps by manipulating the structure of recall and recognition tests. Savings studies manipulating retrieval cues also are overdue.

The third issue concerns a recent, compelling distinction between classes of memory tests. Graf and Schacter (1985) separate explicit tests from implicit tests. Explicit tests require conscious access to memory, with recall and recognition as prototype instances. Implicit tests rely on access to memory, but without the constraint of awareness of that access. Examples of such tests are repetition priming in word fragment completion and in lexical decision. Where does relearning/savings fit into this dichotomy? Intuitively, it would seem to be an implicit test, because savings does not require awareness. Indeed, savings has been demonstrated to occur without awareness (cf. Titchener, 1923; MacLeod, 1976), and the pattern of savings in Table 8 is unlike the pattern of effects we observe in standard recall and recognition studies. Yet the savings pattern also does not seem to correspond to that for implicit tests such as fragment completion, if the results in Table 8 are compared with the findings Roediger and Blaxton (1987) reported, for example. Certainly, this is a topic that warrants further investigation.

Ultimately, the savings paradigm may help us to understand the nature of representation in long-term memory. From Nelson's experiments (Nelson, 1971, 1978; Nelson et al., 1979, 1984; Nelson & Rothbart, 1972), as well as those of Conover and Brown (1977), MacLeod (1976), and the present experiments, we are already gaining some insight into the complexity of this representation. An important goal—one that Nelson has strongly advocated—is the bringing together of results from many different kinds of memory measures to develop a unified view of representation. This enterprise has been avoided for too long in the study of memory.

The findings reported in this article may lead to a revised conception of savings in long-term memory. At the heart of this alternative is the idea that relearning facilitates the retrieval of information, rather than (or perhaps in addition to) increasing its trace strength. It may be because recall and recognition use different retrieval information that savings is not evident on a recognition test of relearning. This is rather a different position than the longstanding one, wherein relearning increases memory strength in an additive fashion. Emphasis now is on retrieval rather than storage. Less central is the other new idea that related items will show savings when experimental conditions permit subjects to include a relation during encoding, but not when that relation must be excluded during encoding. Some tests of both of these ideas have been suggested to illustrate how they might be evaluated.

In the final analysis, relearning can tell us a good deal about memory in broader terms, two cases in point being the relation between various tests of retention and the nature of representation in long-term memory. There is great potential in the study of savings for the increased understanding of what we remember and how we forget. The experiments and hypotheses in this article only begin to scratch the surface.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Bahrick, H. P. (1967). Relearning and the measurement of retention. *Journal of Verbal Learning and Verbal Behavior, 6,* 89–94.

Bahrick, H. P. (1979). Broader methods and narrower theories for memory research: Comments on the papers by Eysenck and Cermak. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 141–156). Hillsdale, NJ: Erlbaum.

Bahrick, H. P., & Bahrick, P. O. (1964). A re-examination of the interrelations among measures of retention. *Quarterly Journal of Experimental Psychology, 16,* 318–324.

Battig, W. F. (1965). Procedural problems in paired-associate learning research. *Psychonomic Monograph Supplements, 1,* 1–12.

Battig, W. F., & Spera, A. (1962). Rated association values of numbers from 0–100. *Journal of Verbal Learning and Verbal Behavior, 1,* 200–202.

Bower, G. H., & Bostrum, A. (1968). Absence of within-list PI and RI in short-term recognition memory. *Psychonomic Science, 10,* 211–212.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and recognition* (pp. 1–35). New York: Wiley.

Bunch, M. E. (1941). The measurement of retention by the relearning method. *Psychological Review, 48,* 450–456.

Burtt, H. E. (1932). An experimental study of early childhood memory. *Journal of Genetic Psychology, 40,* 287–294.

Burtt, H. E. (1941). An experimental study of early childhood memory: Final report. *Journal of Genetic Psychology, 58,* 435–439.

Conover, J. N., & Brown, S. C. (1977). Item strength and input location in free recall learning. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 109–118.

Ebbinghaus, H. (1885). *Über das Gedächtnis* [Memory]. Leipzig, Germany: Duncker and Humblot.

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

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501–518.

Groninger, L. K., & Groninger, L. D. (1980). A comparison of recognition and savings as retrieval measures: A reexamination. *Bulletin of the Psychonomic Society, 15,* 263–266.

Hanawalt, N. G., & Tarr, A. G. (1961). The effect of recall upon recognition. *Journal of Experimental Psychology, 62,* 361–367.

Humphreys, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning and Verbal Behavior, 17,* 175–187.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. *Canadian Journal of Psychology, 36,* 300–324.

Kintsch, W. (1968). Recognition and free recall of organized lists. *Journal of Experimental Psychology, 78,* 481–487.

Kintsch, W. (1977). *Memory and cognition.* New York: Wiley.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 554–565.

Loftus, G. R. (1982). Picture memory methodology. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 257–285). New York: Academic Press.

Luh, C. W. (1922). The conditions of retention. *Psychological Monographs, 31*(3, Whole No. 142).

MacLeod, C. M. (1976). Bilingual episodic memory: Acquisition and forgetting. *Journal of Verbal Learning and Verbal Behavior, 15,* 347–364.

MacLeod, C. M., & Nelson, T. O. (1984). Response latency and response accuracy as measures of memory. *Acta Psychologica, 57,* 215–235.

Mandler, G. (1979). Organization and repetition: Organizational principles with special reference to rote learning. In L.-G. Nilsson (Ed.), *Perspectives on memory research* (pp. 293–327). Hillsdale, NJ: Erlbaum.

Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 386–396.

McGovern, J. B. (1964). Extinction of associations in four transfer paradigms. *Psychological Monographs, 78*(16, Whole No. 593).

Miller, G. A. (1958). The free recall of redundant strings of letters. *Journal of Experimental Psychology, 56,* 485–491.

Muir, B. M. (1982). *An examination of the relationship between the feeling of knowing and savings.* Unpublished senior honors thesis, University of Toronto, Scarborough Campus.

Nelson, T. O. (1971). Savings and forgetting from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 568–576.

Nelson, T. O. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 453–468.

Nelson, T. O. (1985). Ebbinghaus's contribution to the measurement of retention: Savings during relearning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 472–479.

Nelson, T. O., Fehling, M. R., & Moore-Glascock, J. (1979). The nature of semantic savings for items forgotten from long-term memory. *Journal of Experimental Psychology: General, 108,* 225–250.

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282–300.

Nelson, T. O., & Rothbart, R. (1972). Acoustic savings for items forgotten from long-term memory. *Journal of Experimental Psychology, 93,* 357–360.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Postman, L. (1978). Picture-word differences in the acquisition and retention of paired associates. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 146–157.

Postman, L., Jenkins, W. O., & Postman, D. L. (1948). An experimental comparison of active recall and recognition. *American Journal of Psychology, 61,* 511–519.

Postman, L., & Rau, L. (1957). Retention as a function of the method of measurement. *University of California Publications in Psychology, 8,* 217–270.

Postman, L., & Stark, K. (1969). The role of response availability in transfer and interference. *Journal of Experimental Psychology, 79,* 168–177.

Roediger, H. L., III, & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus Centennial Conference* (pp. 349–379). Hillsdale, NJ: Erlbaum.

Schacter, D. L. (1985). Priming of old and new knowledge in amnesic patients and normal subjects. *Annals of the New York Academy of Sciences, 444,* 41–53.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163.

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior, 23,* 3–22.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology, 25,* 207–222.

Titchener, E. B. (1923). Relearning after 46 years. *American Journal of Psychology, 34,* 468–469.

Tulving, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 37–73). London: Wiley.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 479–496.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Tversky, B. (1979). Pictorial representations in adults and children. *Quarterly Journal of Experimental Psychology, 31,* 397–408.

Watkins, M. J., & Todres, A. K. (1978). On the relation between recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 17,* 621–633.

Weldon, M. S., & Roediger, H. L., III (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 15,* 269–280.

Whitten, W. B., Suter, W. N., & Frank, M. L. (1979). Bidirectional synonym ratings of 464 noun pairs. *Journal of Verbal Learning and Verbal Behavior, 18,* 109–127.

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review, 86,* 44–60.

Appendix A

Experiment 1: Materials for Original Learning and Relearning

Stimulus no.	Original learning item	Relearning item
17	MUSHROOM	BOOK
26	GUITAR	CHAIR
28	CIGARETTE	DRUM
30	GLASS	ENVELOPE
32	SHIRT	FORK
35	BUTTERFLY	HAMMER
42	TOASTER	KEY
44	APPLE	PENCIL
48	DOOR	SANDWICH
55	BELT	TREE
56	SNAKE	ASHTRAY
70	SCISSORS	WHISTLE
72	NOSE	BARREL
76	SLED	CANDLE
77	BED	FOOTBALL
80	CHAIN	GLOVE
81	KITE	HORSE
94	RING	LADDER
95	UMBRELLA	PUMPKIN
97	RABBIT	WINDMILL

Appendix B

Experiment 3: Materials for the Identical, Same-Referent, and Control Conditions for Original Learning and Relearning

Stimulus no.	Same-Referent Picture A	Same-Referent Picture B	Control picture
17	PIG	pig	typewriter
23	SAILBOAT	sailboat	flowers
26	CAKE	cake	picture
28	CLOCK	clock	bridge
30	TELEVISION	television	lawnmower
32	FISH	fish	stairs
35	SPEAR	spear	luggage
40	SOCK	sock	car
42	BUTTERFLY	butterfly	frying pan
44	CHAIR	chair	teapot
48	SCALE	scale	purse
55	HAT	hat	tree
56	BARN	barn	refrigerator
63	DRESS	dress	bicycle
70	TENT	tent	camera
72	BARREL	barrel	iron
76	TRUCK	truck	peas
77	BED	bed	streetcar
80	GIRL'S FACE	girl's face	desk
81	HOUSE	house	pipe
89	LION	lion	telephone
94	SCISSORS	scissors	castle
95	LAMP	lamp	umbrella
97	CUP	cup	binoculars

Note. Picture A is in uppercase and Picture B is in lowercase to indicate that these are physically nonidentical pictures referring to the same object.

Appendix C

Experiment 5: Materials for Original Learning and Relearning

Stimulus no.	Response List A	Response List B
17	MUSHROOM	BOOK
23	CROWN	BALLOON
26	CHAIR	GUITAR
28	DRUM	CIGARETTE
30	GLASS	ENVELOPE
32	SHIRT	FORK
35	HAMMER	BUTTERFLY
39	CLOCK	FROG
40	HEART	BROOM
42	TOASTER	KEY
44	APPLE	PENCIL
48	SANDWICH	DOOR
52	ARROW	VEST
55	BELT	TREE
56	SNAKE	ASHTRAY
62	LEMON	THUMB
63	BUS	STAR
70	SCISSORS	WHISTLE
72	NOSE	BARREL
76	SLED	CANDLE
77	BED	FOOTBALL
80	CHAIN	GLOVE
81	KITE	HORSE
85	SOCK	CARROT
89	TOOTHBRUSH	OWL
94	LADDER	RING
95	PUMPKIN	UMBRELLA
97	WINDMILL	RABBIT

Received May 23, 1986
Revision received January 13, 1987
Accepted February 11, 1987 ■