The declining course of recognition memory*

ARTHUR I. SCHULMAN

University of Virginia, Charlottesville, Virginia 22901

Recognition memory for words can decline, sometimes precipitously, over the course of the recognition test. Decrements of as much as 75% in d' have been observed from the first quarter of testing to the last, and decrements of 40%-50% are not uncommon. Proactive mechanisms, imperfectly understood, appear to underlie such declines: Processing early input renders late input less recognizable, and making early recognition judgments renders later judgments more difficult. At the same time, the processing of late-input words fails to inhibit the recognition of the words that preceded them.

There is little in the experimental literature to suggest that the ability to recognize words recently encountered may decline, sometimes precipitously, over the course of a test of recognition memory. And yet, as this paper will show, this is often the case. My purpose here is to present the evidence, for theories of memory must acknowledge it, and to explore, in a tentative way, some of its implications.

THE EVIDENCE TO DATE

The bulk of the evidence is new or unpublished, but the most pronounced declines have already been reported (Schulman, 1971). The latter are reproduced here in Fig. 1, which shows the declining course of recognition memory under a two-alternative forced-choice procedure for words recently classified according to structural or semantic criteria. An unanticipated recognition test followed rapid self-paced decisions as to whether or not each of 96 words contained the letter "A" (A), had no repeated letter (D), denoted a living thing (L), or denoted a geographical location (G). (Under Condition LG each word had to be characterized either as a living thing *or* as a geographical location.) The order in which words were classified was highly correlated with their test order ($r = .93$), so that the lag between input and test was nearly the same for all words. Since the same words were classified and the same test pairs used under all conditions, it may be no surprise to find similar decreasing functions, despite differences in the type of classification required. What is remarkable about the functions of Fig. 1, however, is the extent of their decline. Measured by the discriminability parameter d' , borrowed from signal detection theory (Green & Swets, 1966), recognition fell by 55%-75% from the first quarter of testing to the last.

Previously unpublished functions similar to those of Fig. 1 were obtained with the same recognition test used by Schulman (1971) but following variations in input

procedure. First-to-last-quarter decrements in d' were 58% and 55% (when the words classified at input as L or G included, respectively, an initial or terminal string of 24 that were not to be tested¹); 41% (also as under Condition LG but with forewarning of the recognition test); and 47% (with slide presentation at a 4.5-sec rate, no classification judgments required and, again, with forewarning of the recognition test).

Substantial, though previously unreported, decrements have also been observed in a group of experiments on recognition memory (Schulman, 1967; Schulman & Lovelace, 1970; Schulman, 1973), in which input procedures required no classificatory judgments, words were learned "intentionally" and not "incidentally," and the input and test positions of words were uncorrelated. These experiments also used the method of confidence ratings instead of the forced-choice method. From each operating characteristic that resulted, a value of d_s —an index of

Fig. 1. The declining course of recognition memory. See text for explanation.

*This paper was prepared at the University of Sussex during my tenure as Sesquicentennial Associate of the Center for Advanced Studies of the University of Virginia. I am grateful to both universities. Some of the research described herein was first reported at the 1969 meeting of the Psychonomic Society.

Fig. 2. The declining course of recognition memory. Functions a, c, d, and f describe performance under Conditions RS, CS, RF, and CF, respectively, in Schulman (1970); C(ommon) or R(are) words were presented at F(ast) or S(low) rates. Function b is drawn from the data upon which Schulman (1967) is based. Function e similarly derives from Experiment II of Schulman (1973).

discriminability related to d' and defined as the difference between the z-score coordinates at the point on the operating characteristic that lies on the negative diagonal—was read off. In terms of d_s , recognition memory declined by 36%-50%, as Fig. 2 shows, from the first quarter of testing to the last.

Peixotto (1947) provides further evidence for the declining course of recognition memory. Her Ss inspected 90 nonsense syllables, 30 of which were later to reappear, apparently in scrambled order, among 90 new syllables in a yes/no test of recognition memory. Test performance declined monotonically, with the decrement from the first fifth of testing to the last being roughly equivalent to a 35% drop in d' .

Similar decrements have been found in experiments on recognition memory for words from categorized lists (Schulman, unpublished). Here, 96-word lists were made up of words drawn from two, four, or eight categories. Under these k-category conditions, each word was preceded by one of the k category names that might—and with probability one-half did—apply to it. Ss decided, as quickly as possible, whether or not each word belonged to its questioned category. They were then tested, unexpectedly, on their ability to recognize the words they had just classified. The recognition test was of the two-alternative forced-choice variety, with a word's test position highly and positively correlated with its position in the classification sequence. First- to last-quarter declines in d' were 39%, 30%, and 33% for the two-, four-, and eight-category conditions, respectively.

A few investigators have reported smaller and more irregular decrements in performance over the course of a

test of recognition memory. Underwood and Freund (1970a), in an experiment testing the theory that frequency discrimination dominates recognition performance, found that recognition for 200 words declined significantly, though slightly and nonmonotonically, with stage of testing. The same investigators (Underwood & Freund, 1970b) tested the recognition of 50 briefly presented words by pairing them with words used equally or unequally often in the language. Here again they found slightly but reliably poorer recognition during the second half of testing than during the first. Relatively small stage-of-testing effects have also been noted² both for words recently the objects of classificatory judgments and for words processed in the course of making those judgments.

"Steady state" recognition appears to be the exception, not the rule. Shepard and Chang (1963) and Shepard and Teghtsoonian (1961) both report steady state recognition, but they used a continuous memory task and their stimuli were three-digit numbers, not words. Norman and Waugh (1968) alternated input and test sequences of rapidly presented English monosyllables and found almost no evidence for a within-test decline in recognition memory. It should be noted that the experiments of Shepard and Chang, Shepard and Teghtsoonian, and Norman and Waugh, unlike the experiments previously described, all interleaved inputs and test items. When this is not done—when, as is usually the case, a long input sequence is followed by a long recognition test—it is hard to find data showing performance to be independent of stage of testing.

SOURCES OF DECLINING RECOGNITION MEMORY: AN EXPERIMENT

What, then, can account for the declining course of recognition memory? Both proactive and retroactive "interference" (PI and RI) often have been invoked to explain performance decrements in learning situations (see, most recently, Postman & Underwood, 1973). The role of RI would seem rather unimportant in long-list recognition memory, although Shepard (1967) reports somewhat better recognition for words drawn from the second half of an extremely long inspection sequence than from the first. Yet, the high first-quarter recognition levels of Fig. 1, the result cited in Note 1, and common experience all suggest only minor inhibitory effects of late-input processing upon early-input recognition—an expectation first confirmed by Heine (1914) but whose empirical foundations remain in need of reinforcement. The experiment described below was designed to provide some of the evidence needed to resolve this issue.

Suppose that words from a short list are easier to recognize than words from a long list only because short lists are relatively immune to PI. If so, provided the first words of an input sequence are also the first words

whose recognition is tested, the ability to recognize them should not vary with the number of words that follow them at input. This notion was tested with the help of John A. Merchant. Because this experiment has not previously been reported, its procedure is described in conventional detail below.

Method

An input sequence of 100 common English nouns was selected at random from 200 such words (A and AA frequency, Thorndike & Lorge, 1940, G count), which themselves had been randomly chosen. The 100 words that remained were randomly paired with the input words in a two-alternative forced-choice test of recognition memory. Input words were presented singly on slides at a 2-sec rate that included 0.75 sec change time (a Kodak carousel projector was used). The test of recognition memory was administered in booklet form. 25 pairs to a test page. Each pair was presented on a separate line, with the "old" word as likely to appear on the left as on the right. Ss were told, in instructions given prior to the input sequence, to place a checkmark near the word that had previously appeared on a slide. All Ss were presented with the same initial sequence of 25 input words; these reappeared, in scrambled order, on the first page of the recognition test booklet. The second, third, and fourth quarters of the input sequence were similarly tested on corresponding pages of the recognition booklet.

There were three main experimental groups. One was presented with (and tested on) only the first 25 words of the prepared input sequence, one with the first 50 words of that sequence, and one with the complete 100-word sequence. In order to minimize short-term memory advantages that Groups 25 and 50 might otherwise have had, a filler task was inserted between input and test for all Ss. This task required the circling of all two-digit numbers divisible by 3 and the crossing out of all others. Group 100 was engaged in this task for 100 sec, Group 50 for 200 sec, and Group 25 for 250 sec, so that the time from the start of the input sequence until the start of the recognition test was the same for all Ss. In addition to the three main experimental groups, two groups were run without the intervening task and were tested immediately upon the conclusion of the input sequence. These groups (25I and 50I) were in other ways treated the same as the main groups.

Each group was tested in a single experimental session. Ss were 71 students drawn from introductory classes in psychology at the University of Virginia. Their unpaid participation either fulfilled a course requirement or permitted them to accrue extra course credit. There were 13, 18, and 15 Ss, respectively, in Groups 25, 50, and 100; there were 13 Ss in Group 25I and 12 in Group 50I.

Results

It was expected that Group 100 would show the usual decline in recognition memory over test quarters. At the same time, it was predicted that first-quarter recognition would be the same for Groups 25, 50, and 100 and that second-quarter recognition would be the same for Groups 50 and 100. All of these expectations were confirmed, as Fig. 3 shows. For Group 100, d' declined by about 52% from the first quarter of testing to the last. Neither the second-quarter comparison between Groups 50 and 100 nor any of the paired first-quarter comparisons between Groups 25, 50, and 100 came close to yielding a significant difference in recognition. Figure 3 also indicates that immediate testing resulted in

Fig. 3. Recognition memory as a function of stage of testing for input lists of 25, 50, and 100 words. Words tested for recognition in any quarter were the same for all groups and had appeared in the same input positions for all groups. Groups 25I and 50I were tested immediately after input, while the other groups were tested after brief interpolated activity.

better recognition than testing delayed by the intervening task. The first-quarter difference between Group 25I and Group 25 is reliable ($t = 3.01$, $df = 31$, $p < .01$), as is the second-quarter difference between Groups 50I and 50 ($t = 2.01$, $df = 28$, $p \approx .05$); but the first-quarter difference between the latter groups failed to approach statistical significance ($t = 1.14$, $df = 28$, $p > .20$).

Discussion

The foregoing results support the claim that deleterious proactive effects—of the processing of early input upon the encoding of late and of early recognition judgments upon later ones—are responsible for the declining course of recognition memory. They provide no evidence suggesting that a word's recognition is affected by the number of words following it at input.

The experimental design, however, does not let us gauge the relative importance of the two proactive processes. The only evidence I can present on this issue has been described elsewhere (Schulman, 1971). A group of 15 Ss (Group LGR) was tested under the same conditions as Group LG in Fig. 1, except that the recognition test sequence was reversed. If LG's "forgetting function" owed everything to the proactive effects of making recognition judgments, the particular order chosen for the test sequence should be of no consequence and LGR's decline should have been the same as LG's. On the other hand, if the number of prior inputs was all that mattered, LGR's forgetting function should have been the mirror image of LG's. What actually happened was that LGR's performance over test

quarters was virtually constant—about 86% correct, the same overall level shown by LG—indicating that prior inputs and prior decisions were nearly equally helpful for recognition memory. Procedural changes (different processing demands at input, for example) might well alter the relative roles of input and output proaction, but it seems reasonably clear that both are responsible for the declining course of recognition memory.

WHY PROACTION WORKS

There have been frequent discussions of the role of PI in human learning, but none has needed to deal with the forgetting functions presented here and the theoretical problems they pose. In particular, why should the processing of early input lower the recognizability of later input? And why should early recognition judgments lower the chance that later judgments will be correct?

Consider first the question of input proaction. As the words of an input list are presented, S operates on them in ways over which the E has only limited control. Some experiments (e.g., Schulman, 1971, and Schulman, in preparation) require that the words be treated in specific ways, while others permit greater freedom in the way a word is processed. We may suppose, however, that, like any repetitive task, the encoding of a long list of words becomes progressively more standardized and automatic with practice. That is to say, however S may tend to process verbal inputs, he will do so less variably and more “efficiently” as he works his way through the list. Now, the cost of processing efficiency is the detailed awareness of what one has done. But the encoding of such detail may be important for the memory of an event; indeed, there is reason to believe that memory for an event increases with the ability to recall aspects of the encounter. If this argument is correct, input proaction should be greater after repetitive tasks that allow few processing options and, consequently, lend themselves to automatization than after tasks that preclude uniform processing. While only further research can resolve the issue, the largest declines in recognition memory (Schulman, 1971; see Fig. 1) and some of the smallest (Schulman, in preparation) were in fact observed in experiments having such different processing requirements.

As he gains experience in encoding the words of a list, S paradoxically provides himself (or so I have argued) with increasingly inadequate encodings for later recognition. Input proaction, that is to say, may be seen to result from diminishing information storage. But how does output proaction work? Unfortunately, there are many possibilities. Discrimination between old and new words might become progressively more difficult because the encoding of new test words adds to the data pool at which the recognition process is directed. It seems unlikely, however, that test words would produce such “response competition” when late-input words do

not. Alternatively, the recognition process itself might somehow degrade the traces of words not yet tested, perhaps by obliterating “time tags.” It might also serve to “dislodge” these traces, thereby diminishing their accessibility rather than their adequacy. Yet another possibility is that the problem is not with the trace at all, but rather with the test word and how it is encoded. If successive test words were encoded with decreasing similarity to their original encoding, recognition memory clearly would suffer. The evidence at hand regrettably fails to rule out any of these possible mechanisms of output proaction.

CONCLUSIONS

Recognition memory for words can decline precipitously, as we have seen, over the course of a recognition test. Underlying such a decline are the proactive effects of early upon later processing and of early upon later recognition judgments. At the same time, words processed late in the input sequence seem not to inhibit the recognizability of the words that preceded them. It is not clear why input proaction works, but it is suggested that, as S becomes more skilled in the processing of input words, his lexical encodings become increasingly inadequate bases for recognition memory. Since the reasons for test (or output) proaction are even less clear, both input and test PI deserve further study. The forgetting functions presented here pose a warning to future investigators not to design recognition studies so as to render their interpretation contingent upon unlikely steady state performance. These functions carry a positive message as well as a cautionary one, however, in calling for the systematic study of those input and test conditions that determine the declining course of recognition memory.

REFERENCES

- Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.
- Heine, R. Über Wiedererkennen und ruckwirkende Hemmung. *Zeitschrift der Psychologie*, 1914, 68, 161-236.
- Norman, D. A., & Waugh, N. C. Stimulus and response interference in recognition-memory experiments. *Journal of Experimental Psychology*, 1968, 78, 551-559.
- Peixotto, H. E. Proactive inhibition in the recognition of nonsense syllables. *Journal of Experimental Psychology*, 1947, 37, 81-91.
- Postman, L., & Underwood, B. J. Critical issues in interference theory. *Memory & Cognition*, 1973, 1, 19-40.
- Schulman, A. I. Word length and rarity in recognition memory. *Psychonomic Science*, 1967, 9, 211-212.
- Schulman, A. I. Recognition memory for targets from a scanned word list. *British Journal of Psychology*, 1971, 62, 335-346.
- Schulman, A. I. Recognition memory and the recall of spatial location. *Memory & Cognition*, 1973, 1, 156-160.
- Schulman, A. I., & Lovelace, E. A. Recognition memory for words presented at a slow or rapid rate. *Psychonomic Science*, 1970, 21, 99-100.

- Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, 1967, 6, 156-163.
- Shepard, R. N., & Chang, J. Forced-choice tests of recognition memory under steady-state conditions. *Journal of Verbal Learning & Verbal Behavior*, 1963, 2, 93-101.
- Shepard, R. N., & Teghtsoonian, M. Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, 1961, 62, 302-309.
- Underwood, B. J., & Freund, J. S. Testing effects in the recognition of words. *Journal of Verbal Learning & Verbal Behavior*, 1970a, 9, 117-125.
- Underwood, B. J., & Freund, J. S. Word frequency and short-term recognition memory. *American Journal of Psychology*, 1970b, 83, 343-351.

NOTES

1. It is worth noting that the initial string depressed recognition of the words that followed them, whereas the terminal string failed to depress recognition of the words that preceded them. This issue is discussed at some length later in this paper.

2. Schulman, A. I. Memory for words recently classified. In preparation.

(Received for publication April 11, 1973;
accepted May 8, 1973.)