JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 9, 554–560 (1970)

Further Evidence for a Negative Recency Effect in Free Recall[1]

FERGUS I. M. CRAIK, JOHN M. GARDINER, AND MICHAEL J. WATKINS

Birkbeck College, University of London, London, England

The results of a previous experiment had indicated that the final words in a free-recall list were retrieved least well in a second recall session. This "negative recency effect" was found to hold for reminiscence items and, to some extent, for prior list intrusions in recall. Two further experiments showed that words retrieved in free recall were subsequently recognized less well if they had been presented late in their original input list. There was some evidence that a pattern of primacy and negative recency also held for the recognition of words not retrieved in free recall. It was concluded that terminal words in a free-recall list, although best recalled immediately, are thereafter least available in memory.

The phenomenon of negative recency in free recall was reported by Craik (1970). In that study, 10 lists of words were presented to *S*s for immediate free recall (IFR). After recall of the 10th list, *S*s were given a "final recall" trial in which they were asked to write down, in any order, all the words they could remember from the 10 lists. When words recalled in this final session were allocated to their original input positions, the resulting serial position curve showed a primacy effect, a flat middle portion, and a reliable negative recency effect. By contrast, the serial position curve in IFR took the classic form of a primacy effect, a flat middle portion, and a large recency effect. For words presented at the beginning and in the middle of the list, there was some forgetting between IFR and final recall. The terminal words are of especial interest, however, since they showed the best recall in IFR (the recency effect) but were recalled least well in absolute terms in the final recall session (the negative recency effect). It was argued that the finding of a negative recency effect in final recall is of theoretical importance since it makes "one-process" theories of memory difficult to maintain. At a more descriptive level, the finding implies either that words which are immediately recalled from the end of a list are not "transferred" from some short-term store (STS) to a more permanent long-term store (LTS) (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965) or that *S* has not "manufactured" adequate retrieval cues for these terminal words (Tulving, 1968). Whatever the nature of the mechanism, it seems reasonable to say that if words are not rehearsed, they are not adequately registered in LTS.

The present paper provides further evidence for the negative recency effect. It is shown that the phenomenon holds for reminiscence items and, to some extent, for prior list intrusions in recall, and that the effect is still present when a recognition test is substituted for the final recall test.

EXPERIMENT I

The data presented here represent further analyses of Craik's (1970) study. In that experiment, 20 *S*s were given ten 15-word lists for IFR. Words were presented at a 2-sec rate, and *S*s were given 60 sec for free recall on each trial. Each *S* attended four sessions (on different days) to perform the experiment

[1] Experiment I was carried out while the first author was on a year's leave at the University of Toronto. The study was supported by Grant APA 39 from the National Research Council of Canada to Endel Tulving. John M. Gardiner is in receipt of a postgraduate award from the Science Research Council.

under the four combinations of auditory or visual presentation with spoken or written recall. Thus the pooled data, considered here, comprise 800 subject lists. After the 10 immediate recall trials, *S* was given 5 min to write down as many words as he could from the lists he had just received.

Reminiscence Items

Occasionally *S* would recall a word in the final recall session which he had failed to recall immediately after presentation of the list. Such reminiscence items were rather rare in the present experiment (211 cases out of 12,000 words presented), but it is still of interest to examine their input serial positions. If an item's availability in LTS is accurately reflected by the serial position curve reported earlier—that is, a primacy effect, a flat middle portion, and a negative recency effect—presumably reminiscence items should follow the same pattern. The reminiscence items were thus allocated to their original input positions and expressed as a proportion of the number of unrecalled words at that serial position. Figure 1 thus shows the conditional probability of an item being recalled in the final recall session given that it was not retrieved in IFR. Since the numbers involved in the calculation were small, the serial position curve was smoothed by representing each point as the average of that point, the preceding point, and the succeeding point. This was done for all points except the first and last which are the original scores. Since the shape of the serial position curve is the main point of this paper, all curves were treated in a similar fashion. Any statistical calculations were, of course, carried out on the original scores.

Figure 1 shows that reminiscence items follow the pattern described previously. The first two words are more likely, and the last three words less likely, to occur as reminiscence items than are words from the middle of the list. The general trend of decreasing likelihood of reminiscence with serial position was assessed by correlating the conditional probability scores with input serial position. This procedure gave Spearman's $rho(13) = -0.67$, $p < .01$. It is thus concluded that the trend is reliable.

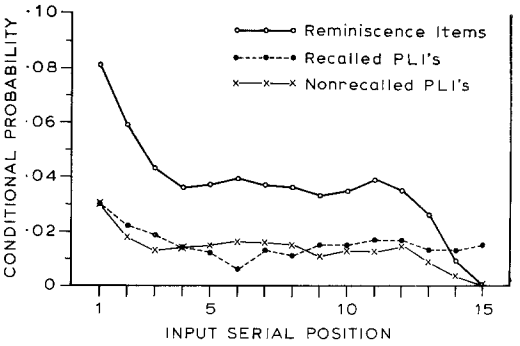

FIG. 1. Conditional probabilities of reminiscence items in final recall, and of two types of prior list intrusion in free recall, as a function of input serial position.

Prior List Intrusions

Welford (1968) cited an unpublished free-recall study by D. A. Kassum who found that intrusions from previously presented lists arose largely from words occupying early input serial positions. Intrusions rarely arose from words presented in terminal serial positions. Welford concluded that these results showed differential registration of words in LTS: the initial words in a presentation list are likely to "capture the translation mechanism and block the entry of subsequent items." His interpretation is thus essentially one of poor transfer of later words from STS to LTS. From Kassum's results and from Craik's (1970) results, it was expected that prior list intrusions (PLI's) would also show the pattern of primacy and negative recency.

The PLI's fall into two categories—items which had been recalled in their correct list and those which had not. These categories might represent somewhat different psychological processes. Accordingly, PLI's were divided into these two classes and allocated to their original input positions. The resulting PLI frequencies were expressed as a proportion of the pool of words from which they

were drawn—that is, the total number of words recalled at each serial position in the first case, and the total not recalled in the second, respectively. Figure 1 thus shows the conditional probability of a PLI given (*a*) that the word had been recalled previously, and (*b*) that the word had not been retrieved in IFR.

Figure 1 shows that PLI's from nonrecalled words followed the predicted pattern of primacy, flat middle portion, and negative recency. When the tendency for these PLI's to arise from initial items was assessed by correlating the conditional probability with input position, using Spearman's technique, *rho*(13) was found to be –0.76, $p < .001$. The PLI's from previously recalled words, however, showed a primacy effect but no negative recency; a similar correlation on this class of PLI was not significant, *rho*(13) = –0.22, $p > .05$.

One further analysis was carried out on the PLI data. Craik (1970) found that words retrieved originally from primary memory (PM) were less likely to be retrieved in the final recall session. A word was categorized as a PM item provided that no more than six items (further stimuli or responses) had intervened between its presentation and recall (Tulving & Colotla, 1970). It was expected that PM items would show a smaller conditional probability of recurrence as PLI's than would SM items. The overall conditional probability of PLI's from PM items was 0.015, and the corresponding conditional probability from SM items was 0.016. These probabilities were not significantly different by the sign test.

The failure to find a difference in the conditional probabilities is somewhat surprising, in view of Craik's (1970) results for correct items and Kassum's results for PLI's. Since the total number of PLI's was small in the present experiment (180 cases out of 12,000 words presented), it is possible that random factors in the data obscured the effect. Accordingly, the data from a previous free-recall study (Craik, 1968; Expt. II) were examined since that study had yielded a greater number of PLI's. In Craik's (1968) experiment, the conditional probability of a PLI from PM items was 0.021, and the corresponding conditional probability from SM items was 0.056. Out of 20 *S*s, 16 showed a higher SM conditional probability, three showed a higher PM probability and there was one tie; the 16/3 split in favor of SM items is significant by the sign test, $p < .01$.

Although not clear, there is thus some evidence to suggest that PLI's are most likely to arise from items presented early in their input list and least likely to arise from items presented late.

Experiment II

Although Craik (1970) found that terminal words in free-recall lists were retrieved least well in the final recall session, it is possible that the terminal words were as well registered in LTS as earlier items but were not retrieved due to less effective retrieval cues (Tulving, 1968). A simple way of testing whether the terminal words were as available in memory but less accessible than words presented earlier would be to follow several free-recall lists by a recognition test. The presence of a negative recency effect under recognition conditions would suggest that terminal words in free-recall lists are poorly registered in LTS.

Procedure

Twenty student *S*s were presented with ten 15-word lists for IFR. The words were drawn from a pool of 230 common, two-syllable concrete nouns. Ten sets of lists were constructed from the same pool and each set was used for two *S*s. The lists were presented orally at a 1-sec rate, and after each list *S* had 60 sec for written recall. The *S*s were tested individually.

After the 10 recall trials, *S*'s response protocols were examined, and words were categorized as retrieved from PM, retrieved from SM, or not recalled (NR). Retrieval from PM was again defined as any case where no more than six words intervened between a word's presentation and recall. For each *S*, an average of three PM items, three SM items and two NR items were selected randomly from each list, to give a total of 80 words. These 80 words were then allocated

at random to positions in a prepared recognition sheet containing the 80 unused items from the common word pool as distractors. Preparation of the recognition test took 15 min on average.

For the recognition test, the 160 words were presented orally at a 4-sec rate. The *S* was informed that half the words would be "old" and that he had to rate each word on a 6-point confidence scale ranging from "certain word is old" (6) to "certain word is new" (1).

Results

Several analyses of *S*s' responses were carried out. First, the rating scale responses were collapsed from 6 to 2 points (either "old" or "new") and percentage recognition scores compared for the different classes of words. Recognition scores were: SM items, 95.8%; PM items, 83.0%; NR items, 57.4%; and distractors (that is, false positives), 30.6%. Wilcoxon matched-pairs signed-ranks tests showed that these sets of percentages were all reliably different from each other at the $p < .001$ level.

A signal detection analysis yielded a similar result. Mean d' scores were as follows: SM items, $d' = 2.60$; PM items, $d' = 1.76$; NR items, $d' = 0.72$. An analysis of variance on the d' scores showed that these classes of words were recognized with different effectiveness, $F(2, 38) = 80.89$, $p < .001$. Again, all comparisons between the means gave reliable differences ($p < .001$) using Scheffé's test.

Finally, the "old" words in the recognition test were allocated to their original serial positions in the free-recall presentation list. Figure 2 shows the mean confidence ratings, given at recognition, both for items recalled and for items which were not recalled in IFR. The data were again smoothed in the manner described in Experiment I. The figure also shows the mean confidence rating given to "new" items in the recognition test.

For the recalled items, the ratings for the first 10 input positions are each based on an average of 44 observations, and the ratings for the last five positions are each based on an average of 134 observations. Figure 2 shows no primacy effect but a marked negative recency effect for the recalled items. The lack of primacy may be due to a ceiling effect. The general tendency for recognition performance to drop with serial position was again assessed by correlating the two variables. This procedure yielded Spearman's $rho(13) = -0.90$, $p < .001$.

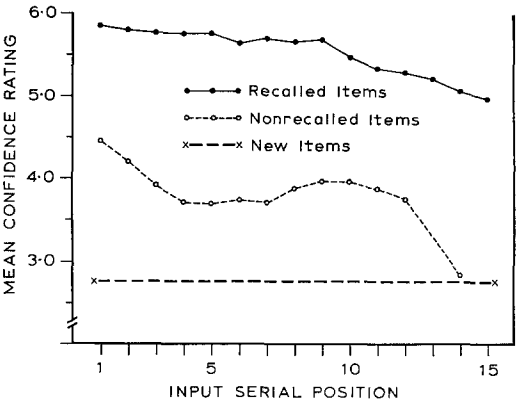

FIG. 2. Recognition performance as a function of input serial position in free recall.

For the nonrecalled (NR) items, the first 12 serial positions are each based on an average of 39 observations, but over the last three positions there were only six observations in all. This was due to the fact that NR items were much rarer at these terminal positions. The last point on the NR curve thus represents the pooled observations from the last three positions. A Spearman rank-order correlation between mean confidence rating and serial position yielded $rho(11) = -0.31$, $p > .05$.

Discussion

The present experiment provides good evidence that the subsequent recognition of words retrieved in free recall is poorer for those words recalled from terminal serial positions. Graphically, the recognition serial position curve for NR items shows a primacy effect and a slight negative recency effect although the tendency for recognition performance to fall with serial position is not statistically reliable. Since the number of observations for the final serial positions was small in the case of

NR items, the results of a further experiment using the same general procedure were examined.

Experiment III

This experiment was one of a series exploring the relationships between items recalled in IFR and in a "final recall" session. In outline, the experiment consisted of 10 IFR trials followed by a final recall trial and then by a recognition test. In the present context, the interest is in the subsequent recognition of words retrieved in one, both, or neither of the recall sessions.

Procedure

Twenty student *S*s were presented with ten 20-word lists for IFR. A pool of 200 common, two-syllable concrete nouns was used to construct the lists. Each *S* received a different set of lists, randomly constructed from the pool. The lists were presented orally at a 2-sec rate, and after each list *S* had 60 sec for spoken recall. Approximately 3–4 min after recall of the 10th list, *S* was given 10 min to write down all the words he could remember from the 10 lists. The *S*s were tested individually.

Immediately after the "final recall" trial, *S* was given a recognition test in which 200 words were presented: 100 words, randomly selected, from the 200 IFR words plus a further 100 distractor words. The 200 words for recognition were presented orally at a 3-sec rate. The *S* rated his confidence that each word was "old" on a 5-point scale ranging from "certain old" (5) to "certain new" (1).

Results

Words recalled in the "final recall" session were recognized extremely well. The mean confidence rating for words recalled both in IFR and final recall was 4.93, and the mean confidence rating for words recalled in final recall but not in IFR (reminiscence items) was also 4.93. Due to ceiling effects, it is thus not possible to plot meaningful serial position curves for these two classes of items.

Figure 3 shows the recognition serial position curves for words recalled in IFR but not in final recall and for words which were not recalled in either session. For the recalled

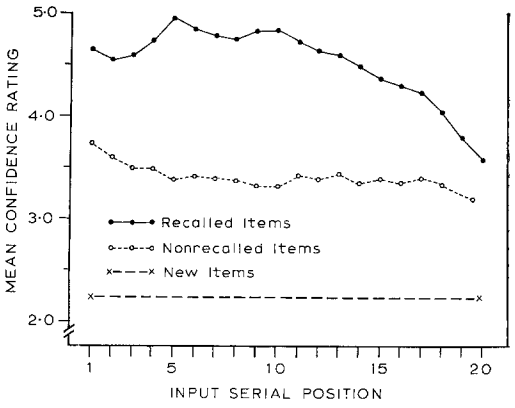

FIG. 3. Recognition performance as a function of input serial position in free recall.

items, the ratings for the first 15 input positions are each based on an average of 12 observations and the ratings for the last 5 positions are each based on an average of 44 observations. The figure shows that recalled items exhibit a strong negative recency effect. A correlation between input serial position and mean confidence rating gave $rho(18) = -0.67$, $p < .01$.

For the NR items, the first 18 serial positions are each based on an average of 50 observations but again there were rather few observations for the last two serial positions. The 24 observations from these two positions were pooled to give a more reliable value for the end of the curve. Graphically, the curve for the NR items shows a slight primacy effect but is thereafter essentially flat. A correlation between mean confidence rating and serial position yielded $rho(17) = -0.52$, $p < .05$. Although the last point in the NR curve shown in Figure 3 has the lowest value, it should be pointed out that the significant correlation probably owes more to the primacy effect than to the negative recency effect.

Discussion

This experiment thus confirms the finding from Experiment II that the recognition of words previously recalled in IFR is poorer for the later input positions. The presence of nega-

tive recency in the serial position curves for NR items is more equivocal and must await further data for clarification.

In general, the results of Experiments II and III support the notion that terminal words in free-recall lists are least well registered in LTS. The presence of negative recency using a recognition test strongly suggests that terminal words are less available in LTS; it is unlikely that they are simply less accessible.

General Discussion

Before discussing the theoretical significance of the results, it should be pointed out that whereas the negative recency effect in Craik's original paper was an *absolute* effect (that is, terminal words were least well retrieved in final recall), the present paper has dealt with conditional probabilities (given that an item was recalled or not recalled in IFR, what was its subsequent probability of recall or recognition). Although negative recency, defined in this way, is not such a striking phenomenon, the interpretation of the effect remains essentially the same.

The results reported in this paper are in line with the conclusion that the last words in a free-recall list, although best recalled immediately, are thereafter least available in memory. This apparent paradox is resolved if two memory processes, PM and SM, are postulated. The PM process may be thought of as a highly accessible, limited capacity store (Atkinson & Shiffrin, 1968) or as the dependence of *S* on effective, but transient, acoustic or temporal retrieval cues (Tulving, 1968). The essence of SM involvement seems to be the integration of new material with the body of past learning; whether this integration is described as "transfer to LTS" (Atkinson & Shiffrin, 1968) or as "construction of semantic retrieval cues" (Tulving, 1968) would seem to be largely a matter of individual taste.

The present results, together with those of Craik (1970), demonstrate that retrieval of a word in immediate recall does not by itself guarantee SM registration. For adequate registration to occur it is apparently necessary to process a word for some time. In the case of terminal words, which are typically output first, this processing time is shorter and registration is correspondingly poorer. The processing time in question is presumably occupied by rehearsal which, after Waugh and Norman (1965), may be thought of as performing two functions: maintaining information (or acoustic retrieval cues) in PM and transferring information (or generating semantic retrieval cues) in SM.

Finally, the relevance of the negative recency effect to two other sets of studies should be pointed out. First, Glanzer and Cunitz (1966) showed that 30 sec of interpolated activity before free recall removed the recency effect. More importantly in the present context, however, their resulting serial position curve showed no negative recency effect. Two comments are offered on this apparent discrepancy with the present results. First, in the Glanzer and Cunitz study, *S*s knew that they still had to recall the words after the delay so they would probably attempt to rehearse each word as it was presented. This contrasts with the present technique in which *S*s output terminal words, thereby fulfilling the main requirement of the task. The second comment is that, in the delayed recall situation, when the interpolated task is very demanding, a slight negative recency effect does occur for the last word (Glanzer, Gianutsos, & Dubin, 1969; Experiments 2, 3, and 5).

The second finding, pertinent to the negative recency effect, is that of retrograde interference in free recall reported by Tulving (1969). He found that when a high-priority item, such as a famous name, was inserted into a free-recall list with the instructions that *S* must remember the name and recall it first, the word immediately preceding the high-priority item was recalled less well than the corresponding word in a control list. Tulving explained the finding in terms of poorer

consolidation of the item preceding a high-priority event. Although the time characteristics of the negative recency effect are somewhat different from the retrograde interference effect, it seems likely that the two effects are manifestations of the same basic phenomenon. In the case of the negative recency effect, this implies that the process of registration in SM continues for a few seconds after the word has been presented and that the process is terminated when the word is recalled.

References

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation*, Vol. 2. New York: Academic Press, 1968. Pp. 89–195.

CRAIK, F. I. M. Two components in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 996–1004.

CRAIK, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 143–148.

GLANZER, M., & CUNITZ, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

GLANZER, M., GIANUTSOS, R., & DUBIN, S. The removal of items from short-term storage. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 435–447.

TULVING, E. Theoretical issues in free recall. In T. R. Dixon and D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968. Pp. 2–36.

TULVING, E. Retrograde amnesia in free recall. *Science*, 1969, **164**, 88–90.

TULVING, E., & COLOTLA, V. Free recall of trilingual lists. *Cognitive Psychology*, 1970, **1**, 86–98.

WAUGH, N. C., & NORMAN, D. A. Primary memory. *Psychological Review*, 1965, **72**, 89–104.

WELFORD, A. T. *Fundamentals of skill*. London: Methuen, 1968. Pp. 214–216.

(Received May 8, 1970)