Retrieval difficulty and subsequent recall*

JOHN M. GARDINER†, FERGUS I. M. CRAIK††, and FRASER A. BLEASDALE†††

Birkbeck College, University of London, London, England

The notion that difficult initial retrieval facilitates subsequent recall was tested in a situation similar to Brown and McNeill's (1966) tip-of-the-tongue (TOT) paradigm. After 50 trials, Ss were unexpectedly asked to recall all the target words. It was found that words retrieved with difficulty in the definition session were relatively well recalled in the final test. Further analyses revealed that the critical factor for good recall was the presence of a TOT state, or a strong feeling of knowing the word, during initial retrieval. An explanation in terms of activation of the word's attributes was suggested.

Craik (1970) described an experiment in which 10 immediate free recall (IFR) trials were followed by a final recall test, in which Ss attempted to recall as many words as they could from all of the preceding lists. Craik found that the last few words from the input list, although best recalled in IFR, were recalled least well in the final recall test. In addition to reporting this negative recency effect, Craik also described an unexpected relationship between output order in IFR and final recall probability—namely, the later an item's output position in IFR, the more likely it was to be retrieved again in the final test. The finding that words retrieved early in IFR had the lowest recall probability in the final test is consistent with the notion that these items were retrieved initially from primary memory (PM), since it is known (Bjork, 1970) that items retrieved from PM are not well registered in long-term or secondary memory (SM). The finding that probability of final recall continued to rise through the later output positions was unexpected, however, since items retrieved towards the end of an IFR trial are presumably retrieved with some difficulty and may be described as having a weak trace strength or as being relatively inaccessible to the retrieval process. On the assumption that all items after the first three of four were retrieved from SM, it was expected either that output position would have no effect on subsequent recall or that the items retrieved earliest from SM would have the highest subsequent recall probability. The second possibility follows from the "spew hypothesis" (Underwood & Schulz, 1960).

One possible explanation of the finding is that difficult initial retrieval may somehow be more beneficial for later retrieval (Craik, 1970). However, several authors have pointed out that not all Ss contribute equally to the later output positions. When the data are corrected for this S selection artifact through Vincentization procedures (Hilgard, 1938), it appears that the output order effect is either caused by or at least heavily contaminated by this artifact (Bjork, 1970; Darley & Murdock, 1971; Rundus, Loftus, & Atkinson, 1970).

Our present interest was in further examining the hypothesis that difficult initial retrieval may be beneficial for later retrieval. Bearing in mind the vulnerability of the output order data to selection artifacts, it was decided to examine the hypothesis in a quite different paradigm. A situation was required in which S is first exposed to a range of retrieval difficulties and is then given a final recall test. Such a range of retrieval difficulties was found in the Brown and McNeill (1966) investigation of the tip-of-the-tongue (TOT) phenomenon. These authors presented a series of dictionary definitions of low-frequency words and asked S to retrieve each word so defined. In the present paper, two studies are described in which the final free recall of target words was examined following a similar definition session.

In operational terms, retrieval difficulty was defined by the latency of retrieval in the definition session. It was postulated that longer latencies reflected low accessibility and thus greater retrieval difficulty of the target words. On each trial, S was presented with a dictionary definition of an uncommon word which he then attempted to retrieve. Following the attempted retrieval of 50 such target words, S was unexpectedly asked to write down all the target words he could remember.

The first experiment carried out using this paradigm was a pilot study and will not be reported in detail. In outline, 20 Ss were each presented with definitions of 50 words whose frequency count was less than 1 per million (Thorndike & Lorge, 1944). Each S was tested

individually and received the same 50 definitions in a unique random order. The S was informed that speed of retrieval was the point of the experiment, and E recorded each retrieval latency from a stopwatch. After 90 sec had elapsed, or when S successfully produced the target word, E proceeded to the next definition. The Ss were not provided with target words they had failed to retrieve. One minute after the final trial, S was given 10 min for the final free recall of all target words.

Words retrieved in the definition session were split into three categories on the basis of retrieval latency: 0-2, 2-10, and 10-90 sec. Over all 20 Ss, the numbers of words in these categories were 426, 165, and 104, respectively. The final recall probabilities associated with the categories were 0.38, 0.44, and 0.67, respectively. Thus, longer latencies of initial retrieval were associated with higher final recall probabilities. This trend was assessed by the nonparametric Page's L test after computing probabilities of final recall for the three latency periods for each S separately. This procedure yielded $L(3,20) = 226.0$, $p < .001$, and it may thus be concluded that words retrieved with greatest difficulty—as presently defined—were best recalled in the final session.

In addition to the procedure described above, Ss were asked to report TOT states while retrieving the target words in the definition session. In this study, 58 words were retrieved after. TOT reports and the final recall probability associated with this group of words was 0.76. Since TOT words were necessarily confined to the longer latency periods, it seemed possible that the TOT condition alone had led to the high final recall probabilities. Accordingly, TOT words were removed from the data and the final recall probability recalculated for all remaining words retrieved between 2 and 90 sec. This procedure yielded a probability of 0.46, which, although close to the probability for words retrieved in less than 2 sec (0.38), was still reliably greater by Wilcoxon test $[T(20) = 34.5, p < .01]$.

In summary, this first study provided evidence that difficult initial retrieval is followed by good subsequent recall. There was also some evidence, however, that high recall probabilities were associated with words which Ss retrieved after experiencing a TOT state. The main experiment reported in this paper was thus undertaken as a replication and extension of the first study. Further evidence was gathered on "feeling-of-knowing" states (Hart, 1967) during initial retrieval and on the extent to which these states predicted subsequent recall.

METHOD

The Ss were 30 volunteers, mostly undergraduate students, who were either paid for their services or were fulfilling a course requirement.

Dictionary definitions of 50 words whose frequency counts were less than 1 per million (Thorndike & Lorge, 1944) were written on separate library cards. Again, each S was tested individually and received the same 50 definitions in a unique random order.

Each trial consisted of E reading out a definition plus the initial letter of the target words. The initial letter was given in order to reduce ambiguity. The S was instructed to respond by saying the target word as rapidly as possible after each definition had been presented. The S was given 60 sec to continue his retrieval attempt when he could not retrieve the word immediately. At the end of each 15-sec period, S was required to indicate how close he felt he was to retrieving the target word. Each period was signaled by a metronome, and S wrote down 0 ("no feeling-of-knowing the word"), 1 ("slight feeling-of-knowing the word"), 2 ("strong feeling-of-knowing the word"), or T ("a TOT state"). The S continued to give such ratings either until he succeeded in retrieving the word or until 60 sec had elapsed, at which time the word was read out by E. In cases where E supplied the word, S also indicated whether it was indeed the word he had been trying to retrieve. If S retrieved the word before 60 sec had elapsed, the next definition was presented immediately. The S was instructed to construe the "feeling-of-knowing" ratings as representing the likelihood of his being able to correctly identify the word in a recognition test.

Approximately 1 min after the final trial, S was unexpectedly asked to write down, in any order, all the target words he could remember. He was allowed 10 min for this final recall test.

RESULTS

In the definition session, a total of 776 words were retrieved by all 30 Ss and 724 were supplied by E. Considering first the 776 retrieved words, 634 were retrieved in the first 15-sec period and the remaining 142 were retrieved between 15 and 60 sec. The conditional probabilities of final recall associated with these two classes of words were 0.27 and 0.48, respectively. For the purposes of statistical evaluation, the data supplied by each block of three, consecutively tested Ss were pooled to create 10 macro-Ss (Bregman, 1968; Bruce & Papay, 1970). This was done since not all Ss contributed data to the cells of this and subsequent comparisons—it should be noted, however, that the pattern of results obtained does not depend critically upon this form of data treatment. All 10 macro-Ss showed higher final recall of words retrieved initially between 15 and 60 sec compared with words retrieved initially in less than 15 sec ($p < .01$, by the sign test). Thus, the words which Ss found difficult to retrieve after hearing the definition were recalled significantly better in the final test.

The 142 words retrieved between 15 and 60 sec comprise 49 items which Ss rated 0 or 1 (described here for convenience as "non-TOT items") and 93 items rated 2 or T ("TOT items"). The probabilities of final recall for non-TOT and TOT items were 0.27 and 0.59, respectively. These probabilities were found to be reliably different in a macro-S analysis [Wilcoxon $T(10) = 4$, $p < .01$]. It will be noted that the probability of final recall of non-TOT items retrieved between 15 and 60 sec was identical to that for items retrieved in less than 15 sec (before the first rating was asked for). That is, the benefit conferred to subsequent recall performance by longer retrieval latency was confined to TOT items.

The 724 words which E initially supplied included 127 words which Ss failed to recognize as those they had

been trying to retrieve. These 127 items had a final recall probability of 0.22; they were excluded from further analysis. Of the remaining 597 words, 460 were rated 0 or 1 (non-TOT words) and 137 were rated 2 or T (TOT words). The final recall probabilities of these two classes of words were 0.36 and 0.49, respectively. The superior recall of TOT items was shown by all 10 macro-Ss. Thus, if S thought he was close to retrieving a word in the definition session (although he did not, in fact, retrieve it), the word was more likely to be recalled later. It should be pointed out that in these analyses, if S changed his rating for a particular target word across the 15-sec intervals, the rating used in the analysis was the final rating given.

Two further results may be mentioned briefly. First, final recall was not affected reliably by retrieval or nonretrieval of a word in the initial definition session. For both TOT words and non-TOT words, those items supplied by E were as well recalled as those which had been successfully retrieved by S. For TOT items, the final recall probabilities for retrieved and nonretrieved words were 0.59 and 0.49, respectively; for non-TOT items, the probabilities were 0.27 and 0.36, respectively. In both cases, Wilcoxon tests yielded $T(10) = 14$, $p > .05$. Second, final recall probability was not affected by the duration of the reported TOT states.

DISCUSSION

These slightly complex results are best described in two stages. First, there is good evidence that words retrieved "with difficulty" from their definitions are better remembered than are words retrieved easily. Both the pilot study and the main experiment show this result clearly.

The second stage analyzes this phenomenon at a deeper level. While it was originally believed that the latency of initial retrieval determined final recall, the results described above make it clear that the critical factor is not latency, but rather the presence of a strong feeling of knowing or TOT state during retrieval. Thus, the three classes of non-TOT words—words retrieved rapidly after the definition was presented, words retrieved after at least 15 sec of search but retrieved suddenly after S had reported little feeling of knowing the word, and third, words supplied by E after S had reported little feeling of knowing—were all recalled poorly in final recall (0.27, 0.27, and 0.36, respectively). On the other hand, words which gave rise to TOT states were relatively well recalled in final recall regardless of whether they had initially been retrieved by S or supplied by E (0.59 and 0.49, respectively). Good recall did not depend on successful retrieval in the definition session. Rather, good recall followed when S felt himself to be close to retrieval in the definition session and reported the occurrence of a strong feeling of knowing or TOT state.

The good recall of TOT words was also found in the pilot study, although an effect of retrieval latency remained when such words were excluded. This anomalous result may be due to the fact that Ss only reported words which were "on the tip of the tongue." Strong "feeling-of-knowing" words may have remained in the analysis and contributed to the higher recall probability for the longer retrieval intervals.

Why should words with TOT states be well remembered? Two possibilities may be suggested. The first is that when the desired word is finally retrieved or supplied, it receives more attention than do non-TOT items. The superior recall of TOT items may thus be similar to the von Restorff effect. Alternatively, it seems likely that while S is attempting to retrieve a word, some of the word's attributes or associates may be activated. In this case, S will hold these attributes in "working memory" in an attempt to generate the target (Brown & McNeill, 1966) and thereby give rise to a feeling of knowing the word or to a TOT state. It is known that the activation of a word's semantic attributes leads to good recall, even in an incidental learning paradigm (Hyde & Jenkins, 1969). By this explanation, the critical difference between words retrieved immediately and those retrieved later is that some members of the latter group pass through a stage in which their attributes and associates are activated and processed. Although we consider the second alternative more attractive, the explanations cannot be evaluated on the basis of the present data.

Two concluding points may be made about the results. First, the finding that words retrieved easily in the definition session were least well recalled subsequently is analogous to the phenomenon of negative recency (Craik, 1970). Craik found that the last few words in a list, although recalled best in IFR, were recalled least well in the final recall test. Both the present results and the negative recency effect may depend on the fact that a word's semantic-associative attributes must be activated for subsequent recall to be successful. It does not seem to matter whether they are activated by intentional learning, by an appropriate incidental task, or by the process of partially successful retrieval.

Second, the finding that words retrieved easily in the definition session were least well recalled subsequently indicates that the two retrieval tasks differ in some crucial way. One way to describe this difference is that, whereas the definition session involves retrieval from semantic memory, the subsequent recall test involves retrieval from episodic memory (Tulving, 1972). The present findings lend further support to the distinction between episodic and semantic memory.

REFERENCES

Bjork, R. A. Control processes and serial position effects in free recall. Paper presented at the Mathematical Psychology Meetings, Miami Beach, 1970.

Bregman, A. S. Forgetting curves with semantic, phonetic, graphic and contiguity cues. Journal of Experimental Psychology, 1968, 78, 539-546.

Brown, R., & McNeill, D. The 'tip-of-the-tongue' phenomenon. Journal of Verbal Learning & Verbal Behavior, 1966, 5, 325-337.

Bruce, D., & Papay, J. P. Primacy effect in single-trial free recall. Journal of Verbal Learning & Verbal Behavior, 1970, 9, 473-486.

Craik, F. I. M. The fate of primary memory items in free recall. Journal of Verbal Learning & Verbal Behavior, 1970, 9, 143-148.

Darley, G. F., & Murdock, B. B. Effects of prior free recall testing on final recall and recognition. Journal of Experimental Psychology, 1971, 91, 66-73.

Hart, T. J. Memory and the memory-monitoring process. Journal of Verbal Learning & Verbal Behavior, 1967, 6, 685-691.

Hilgard, E. R. A summary and evaluation of alternative procedures for the construction of Vincent curves. Psychological Bulletin, 1938, 35, 282-297.

Hyde, T. S., & Jenkins, J. J. Differential effects of incidental tasks on the organization of recall of a list of highly associated words. Journal of Experimental Psychology, 1969, 82, 472-481.

Rundus, D., Loftus, G. R., & Atkinson, R. C. Immediate free recall and three-week delayed recognition. Journal of Verbal Learning & Verbal Behavior, 1970, 9, 684-688.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Tulving, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972. Pp. 381-403.

Underwood, B. J., & Schulz, R. W. Meaningfulness and verbal learning. Philadelphia: Lippincott, 1960.