

Chunks as Interference Units in Free Recall

GORDON H. BOWER¹

Stanford University, Stanford, California 94305

The experiment tested the hypothesis that free recall is limited by the number of chunks that can be reproduced, not the number of words. Recall of a given word should be higher the fewer chunks are required to encompass the remaining words in the list. To test this, a subset of 12 critical words was embedded in a larger list containing 12 fillers composed either of single words, or three-word cliches, or unrelated word triplets. As predicted, recall of the critical words was the same in the single-word and cliche lists, but poorer in the triplet list. Cliches and single words were equally recalled. Total units recalled in the three lists were the same, with only the selection of units differing in recall. The results confirm the chunking hypothesis and have implications for Murdock's (1960) total-time hypothesis about free recall.

This experiment is concerned with the "chunking" hypothesis applied to free verbal recall (cf. Miller, 1956; Tulving, 1968). For present purposes, a chunk of material may be identified as a highly integrated group of words, indexed by a strong tendency for Ss to recall the words together as a unit. The chunking hypothesis asserts that free recall is limited by the number of chunks that can be produced from memory without the aid of some specific retrieval scheme or cuing system such as the pegword mnemonic (cf. Wood, 1967). According to this hypothesis, recall improves over practice trials because the words become more strongly bonded into subjective chunks, and several chunks may be coalesced into a single chunk by a recursive (hierarchical) process of grouping subgroups. The number of chunks recalled is presumed to remain approximately constant, while the number of experimental items per chunk presumably increases with practice. Papers by Tulving (1968) and Mandler (1967) may be read for a review of evidence relevant to this hypothesis.

¹ This research was supported by Grant MH-13950, from the National Institutes of Mental Health. Bruno Breitmeyer assisted in collecting the data.

The hypothesis implies that a list of N words will be better recalled the fewer are the number of pre-established chunks encompassing the whole list. For instance, a list of six independent words (i.e., six chunks) will not be recalled so well as a list of six words composed of three independent words plus a three-word cliche (i.e., four chunks). In particular, for a given independent word, its recall probability should be higher the fewer are the chunks encompassing the remaining N-1 words of the list. Therefore, the recallability of a given word should depend on the number of other chunks to be recalled, not the number of other words to be recalled.

Prior evidence relevant to this implication is largely indirect, demonstrating that list-conditions differing in words recalled may nonetheless be equal in "chunks" recalled, for some operational specification of chunks in the recall protocol. An example is an experiment by Tulving and Patterson (1968) that compared recall of N unrelated words to recall a list of N-4 unrelated words plus four related words from a distinctive category. More words were recalled from the latter list but this was in large part due to enhanced recall of the related words. If the categorized words

recalled were counted as a single chunk and totaled with the unrelated words recalled, then the number of chunks recalled was nearly constant for the two lists. The evidence is indirect because our hypothesis requires a comparison of recall of the N-4 unrelated words in the two lists, since these occur with one additional chunk in the category list but with four additional chunks in the unrelated list. But the published data are not reported in a form permitting easy computation of the relevant statistics; further, the prior experiments have not been designed to manipulate this variable of interest over an extensive range. For such reasons, the following elementary experiment was carried out to test this implication of the chunking hypothesis.

METHOD

Design

The design involved a within-group comparison of three free recall lists of 24 units. Each 24-unit list consisted of a set of 12 "critical" units (words) plus 12 "filler" units. The filler units were varied over the three lists, and the hypothesis expects concomitant variation in recallability of the 12 critical units in the list. For the one-word list, the fillers were simply 12 nouns presented singly; for the three-word list, the fillers were 12 triplets of unrelated nouns presented as units; for the cliché list, the fillers were 12 familiar, three-word clichés presented as units. The critical words and filler units of a list were mixed for random presentation, and *S* freely recalled all he could. The chunking hypothesis expects recall of the 12 critical words to be about the same for the one-word and the cliché fillers, but significantly poorer with the three-word fillers.

Materials and Procedure

The three sets of 12 critical words were unrelated nouns selected for high concreteness from the norms of Paivio, Yuille, and Madigan (1968). The one-word and three-word fillers were similarly selected by the same criteria. The 12 three-word fillers were composed of unrelated nouns. The 12 clichés were noun phrases of high frequency (intuitive estimates) in our *Ss'* linguistic community. For eight of the clichés, all three words had a noun form, although in the cliché context the first two functioned as adjectives. These eight clichés were: ball-point pen, mail-order catalog, Rose Bowl parade, birth control pill, ice cream cone, Bay Area transit, tick-tack-toe, and turtle neck sweater.

The other four clichés contained some adjectives and were: Happy New Year, fair-weather friend, Great Salt Lake, and good old days. Most of the clichés have a single semantic referent, whereas the unrelated three-word fillers (e.g., couch, flag, sun) have three separate referents.

The three free recall lists were composed by pairing the three critical-word sets with the three filler sets, using the six possible pairing about equally often over *Ss*. The 12 critical words and 12 filler units were typed in capital letters on 24 flash cards, shuffled thoroughly, and shown to *S* at a rate of one card for 3 sec, with a 2-sec intercard interval. There was one study trial, then an immediate recall trial on each of the three lists. The *Ss* recalled in writing, giving as many words as they could in any order. Time allowed for recall was 96 sec for the one-word filler list, and 192 sec for the cliché and three-word filler lists; this is calculated at 4 sec recall time per word on the input list. This was always more than enough time for *Ss* to recall all they could.

The order of the three treatment lists within the session was counterbalanced over *Ss*. The *Ss* were 16 undergraduates fulfilling a service requirement for their introductory psychology course.

RESULTS

An initial observation is that the three-word clichés were recalled in perfect all-or-none fashion, either completely or not at all. On the other hand, the unrelated triplets were not recalled all-or-none; *Ss* frequently recalled some but not all of such triplets (more details will be given later). Therefore, in the following, we adopt the convention of treating a three-word cliché as a single recall unit (i.e., 12 filler units) but unrelated triplets as three separate recall units (i.e., 36 filler units).

In these terms, Table 1 shows the average recall of critical words and of filler units for the three lists. Analysis of variance was performed on the critical-word recall scores after taking arc sine transforms of the percentages (out of 12). The overall *F*-test for treatments is highly significant, $F(2, 30) = 26.3$, $p < .001$. Intergroup comparisons were carried out by the Newman-Keuls procedure, which detected significant gaps between the three-word filler list vs. the one-word and cliché filler lists ($p < .01$ in each case) but an insignificant difference between the one-word and cliché filler

TABLE 1
RECALL OF CRITICAL NOUNS AND FILLER UNITS FOR
THE THREE TYPES OF LISTS

Type	Fillers		
	One-word	Cliches	Three-word
Critical nouns	6.3	5.5	3.7
Filler units	6.7	6.6	9.4
Total units	13.0	12.1	13.1

lists. These results support the implication of the chunking hypothesis: recall of the critical words was about the same whether the other 12 units were one-word or three-word (cliche) units, but recall was reduced if the other words comprised more (than 12) chunks.

The slightly lower recall of the critical words in the cliche list compared to the one-word lists may be attributable to a tendency for *Ss* to recall the cliches first, delaying recall of the critical words in the cliche list. Recall protocols on the cliche list were divided into a first vs. a second half for each *S*, and the number of cliches recalled in each half was tabulated and compared. If no recall-order bias favored the cliches, then approximately half of those recalled should appear in the first half of each protocol. Ten *Ss* had more cliches at the beginning of their recall, four had more at the end, and two showed no difference. A split of 10/4 or more extreme has a binomial probability of .09 on the null hypothesis of no order bias. Therefore, the tendency to recall cliches earlier in the cliche list is not very marked. But it may suffice to account for the small difference in critical-word recall between the one-word vs. cliche filler lists.

Recall of the filler units also supported the chunking hypothesis (line 2 of Table 1). Consistent with hypothesis, the cliches and one-word fillers were equally well recalled. Whether they are recalled better or worse than the 36 unrelated fillers depends on whether frequencies or proportions are considered. As indicated earlier, the unrelated triplets were

not recalled in all-or-none fashion. Of the 192 observations (16 *Ss* × 12 triplets), the frequencies of *Ss* recalling zero, one, two, or three words of the unrelated triplets were 114, 29, 26, and 23, respectively. In 85% of the cases when two or more words were recalled from a given triplet, they came out in adjacent temporal order. This happened particularly with triplets near the end of the input phase, so output adjacency need not suggest unitization of the terms but only simultaneous residency in short-term memory. The mean number of unrelated triplets represented in recall by one or more elements averaged 4.9, which is significantly less than the 6.6 cliche units in recall, $t(15) = 2.50, p < .05$.

A further observation favorable to the chunking hypothesis appears in line 3 of Table 1, giving the total number of recalled units (as previously defined). The total units recalled appear to be the same over the three conditions, $F(2, 30) = 1.25, p > .25$. The equality of total recall in the three cases agrees with Murdock's (1960) empirical generalization that total units in recall increase with the total presentation time for the set of units. This generalization applies to the one-word vs. three-word filler lists because they required the same total presentation time. Although total recall conforms to Murdock's generalization, recall scores from the separate subsets of the lists do not. Despite equality of presentation times within the subsets, more filler words were recalled in the three-word than in the one-word lists, $t(15) = 3.61, p < .001$; but fewer critical words were recalled in the three-word than in the one-word lists. In fact, confirmation of the chunking hypothesis depended upon this latter discrepancy from Murdock's generalization as it would apply to the subset of critical words.

In summary, the results support the chunking hypothesis. In almost every respect, pre-established word groups (cliches) behave like single words in recall, in terms of their recall or in terms of their effect on the recallability

of other units from memory. Recall limits are expressible in chunks, not words. As Tulving and Patterson (1968) concluded: "Thus, the number of functional units of memory that can be retrieved under the conditions of free recall seems to be independent of the size of the units. Recall of constituent elements of a higher-order functional unit appears to occur without any cost to the capacity to retrieve independent units as such."

REFERENCES

- MANDLER, G. Organization and memory. In K. Spence & J. Spence, (Eds.), *The psychology of learning and motivation*, Vol. 1; New York: Academic Press, 1967.
- MILLER, G. A. Human memory and the storage of information. *IRE Transactions on Information Theory*, 1956, 2, 129-137.
- MURDOCK, B. B., JR. The immediate retention of unrelated words. *Journal of Experimental Psychology*, 1960, 60, 222-234.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concrete-ness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, 76, No. 1, Part 2.
- TULVING, E. Theoretical issues in free recall. In T. Dixon and D. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.
- TULVING, E., & PATTERSON, R. D. Functional units and retrieval processes in free recall. *Journal of Experimental Psychology*, 1968, 77, 239-248.
- WOOD, G. Mnemonic systems in recall. *Journal of Educational Psychology Monograph*, 1967, 58, No. 6, Part 2.

(Received April 24, 1969)