Research Article

Constant Capacity in an Immediate Serial-Recall Task

A Logical Sequel to Miller (1956)

Nelson Cowan, Zhijian Chen, and Jeffrey N. Rouder

University of Missouri, Columbia

ABSTRACT—***We assessed a hypothesis that working memory capacity should include a constant number of separate mental units, or chunks (cf. Miller, 1956). Because of the practical difficulty of measuring chunks, this hypothesis has not been tested previously, despite wide attention to Miller's article. We used a training procedure to manipulate the strength of associations between pairs of words to be included in an immediate serial-recall task. Although the amount of training on associations clearly increased the availability of two-item chunks and therefore the number of items correct in list recall, the number of total chunks recalled (singletons plus two-word chunks) appeared to remain approximately constant across association strengths, supporting a hypothesis of constant capacity.***

Miller's (1956) "The Magical Number Seven, Plus or Minus Two" is among the most widely cited works in psychology. Yet it has not spawned much research on a key hypothesis, that immediate memory involves a constant-capacity storage mechanism. One can gather from Miller's article that short-term memory has a capacity of about seven items in immediate-recall tasks, and also that related material can be chunked together to form a new, conjoint item. An example of chunking is that memory for the letter series *IRSCIAFBI* becomes much easier if one recognizes within it three acronyms of U.S. government agencies, IRS, CIA, and FBI. Each acronym becomes a single chunk. However, researchers have not assessed what we term a *constant-capacity hypothesis*, that the number of chunks that can be held in immediate memory is constant no matter what the content of each chunk. Learning associations between items increases the sizes of chunks held in memory and, therefore, the total number of items recalled, but the hypothesis states that this learning does not increase the number of chunks recalled. Miller also discussed a capacity limit in unidimensional absolute identification.

Our main focus in the present study was testing the constant-capacity hypothesis in immediate recall. That hypothesis is not really addressed by Miller's (1956) magic number because each of seven items remembered in an immediate memory task may not be a separate chunk. The seven-item limit is a description of empirical evidence and therefore is not open to much debate, although it can vary when factors such as word length and phonological similarity are manipulated (Baddeley, 1986). The importance of chunking and grouping is almost universally accepted and has been explored in depth (e.g., Anderson, Bothell, Lebiere, & Matessa, 1998; Bowles & Healy, 2003; Ericsson, Chase, & Faloon, 1980; Frankish, 1985; Frick, 1989; Gobet et al., 2001; Hitch, Burgess, Towse, & Culpin, 1996; Marmurek & Johnson, 1978; Ng & Maybery, 2002; Ryan, 1969; Slak, 1970; Towse, Hitch, & Skeates, 1999; Wickelgren, 1967). Unlike these topics, though, the constant-capacity hypothesis has rarely been tested (though see Tulving & Patkau, 1962).

There may be practical as well as historical reasons why the constant-capacity hypothesis has remained untested. Practically, it is difficult to tell exactly how chunking is being used, and therefore how many chunks are being recalled. For example, the finding that people consistently can recall about seven items could occur only because items can be transformed into a smaller number of chunks, which may be formed on an ad hoc and idiosyncratic basis. Indeed, seven-digit telephone numbers are presented in smaller clusters for just that reason. Miller (1956) himself wrote in a way that may have denigrated the constant-capacity hypothesis. He did not explicitly state it, and, furthermore, he concluded with the thought that finding memory for seven items across several procedures may have just been a "pernicious, Pythagorean coincidence" (p. 96), so a sophisticated reading of his article would induce skepticism.

Several researchers have proposed a form of the constant-capacity hypothesis, but with capacity in the range of three to five chunks (e.g., Broadbent, 1975; Mandler, 1985). Cowan (2001) developed this point further by reviewing various experimental situations in which it seemed unlikely that participants could carry out chunking processes. Two out of many examples are array-comparison procedures in which two briefly presented arrays of colored squares are to be compared (Luck & Vogel, 1997) and running-span procedures in which a list of digits of an unpredictable length is presented quickly and recall of items at the end of the list is requested (Pollack, Johnson, & Knaff, 1959). In these cases, and many others in which the items occur too quickly for rehearsal or grouping processes to contribute (see Hockey, 1973), adults retain an average of three to five items from the set. Of

Address correspondence to Nelson Cowan, Department of Psychological Sciences, University of Missouri, 18 McAlester Hall, Columbia, MO 65211; e-mail: cowann@missouri.edu.

course, such findings cannot address the broader hypothesis that the capacity, in chunks, remains constant even if chunking is permitted. Our aim in the present study was to examine that broader hypothesis, the constant-capacity hypothesis.

We circumvented the problem of determining chunking by presenting novel pairs of words (A-B, C-D, E-F, etc.) and examining effects of these paired associations on subsequent memory tests (serial and cued recall). In serial recall, the previously studied pairs were embedded within eight-item lists, and subjects were asked to recall each list immediately after it was presented. If a participant who had encountered an A-B pair recalled the pair in adjacent positions in the list, in order, the pair may have been combined to form one chunk. An alternative possibility is that A and B were recalled separately in adjacent positions; but, as we demonstrate, the probability of that happening can be estimated from recalls of A and B out of order or separated by other words.

Cued recall of the second word in a pair, cued by the first, provides auxiliary information about the long-term representation of word pairs. If cued recall fails for words previously recalled within a list, it can be reasonably surmised that the success in list recall was not based on a permanent chunk in memory. Instead, it might have been based on a pair of words not chunked together or on a temporarily formed chunk.

We presented words as singletons and in pairs and varied the frequency of occurrence of the pairs (one, two, or four paired presentations during a training period). In our data, frequency of presentation of a word pair affected the number of items correctly recalled in both list recall and cued recall. However, testing the effect of frequency of presentation was orthogonal to testing the constant-capacity hypothesis. For the latter test, the critical question was whether the *chunk span*, the total number of two-item chunks and singletons recalled, was constant across training conditions. Finding a constant chunk span would support the constant-capacity hypothesis. We used a simple multinomial model of performance to assist in the theoretically complex task of identifying the number of chunks in recalled lists. Finally, to understand the serial-recall results better, we considered the important issue of the relation between item and order information (cf. Bjork & Healy, 1974; Cunningham, Marmie, & Healy, 1998; Detterman, 1977; Healy, 1974; Lee & Estes, 1981; Lewandowsky & Murdock, 1989; Murdock, 1976; Nairne, Whiteman, & Kelly, 1999; Saint-Aubin & Poirier, 1999).

METHOD

Participants

Thirty-two undergraduates (11 male, 21 female), 16 per experiment, received course credit for participation. They were native speakers of English with no known hearing deficits. Two additional participants were omitted because they did not follow directions.

Stimuli

We selected 198 nouns from the MRC Psycholinguistic Database (Wilson, 1987). Each word had three to four letters, three to five phonemes, one syllable, a Kucera-Francis written word frequency above 12, and a concreteness rating above 500. The words were assigned to three sets by starting with the first word in alphabetical order and cycling among the sets while continuing through the list of words, so each set included 66 words. Different word sets were used for the three trial blocks. For each participant, the stimuli were 40 words randomly selected from each set. These words were presented in 0.64-cm black lettering on white and were viewed at a distance of about 50 cm.

Procedure

Overview

Each participant was tested individually in a quiet booth and performed three experimental blocks. In Experiment 1, each block consisted of a training phase, followed by list recall and then cued recall, whereas in Experiment 2, training was followed by cued recall and then list recall. In both experiments, the main manipulation concerned the presentation method in the training phase.

During training, words were presented singly and in pairs, randomly separated by other words and pairs. Each word was assigned to the *0-*, *1-*, *2-*, or *4-pairings* condition; in each condition, words were presented in consistent pairs the indicated number of times. However, the same words could be presented singly, and in all of these conditions, each word was presented a total of four times, singly or in a pair. Words assigned to a *no-study* condition did not appear in the training phase but were tested subsequently. If a word was presented in a pair, it was presented in the same pair throughout the experiment. Table 1 illustrates the design for Experiment 1, and further details follow.

Training

In the training phase, each word or word pair was presented in the center of the computer screen for 2 s. The participant pronounced the words aloud as they appeared. Altogether, there were 100 single-word and word-pair presentations (16, 24, 28, and 32 presentations in the 4-, 2-, 1-, and 0-pairings conditions, respectively), allowing each word to be presented four times, alone or within a pair.

List Recall

In this phase, five lists were presented, and immediate serial recall was required after each list. Each list comprised eight words, presented in two-word pairs; all words in the same list were from the same training condition, and for words in the 1-pairing, 2-pairings, and 4-pairings conditions, the pairings were the same as in the training phase. The orders of lists and of word pairs within lists were randomized.

The participant initiated each trial. A 1-s waiting period preceded the appearance of the first word pair in the list. Each of the four pairs within the list was presented for 2 s in the center of the screen, with each successive pair replacing the previous one. After the last word pair, the recall test began. The participant was to type all of the words in the presented order. The spelling of each word could be corrected until a space bar was pressed, eliciting a cue to recall the next word. If the participant did not know a word, it was permissible to skip to the next one. The words in the response were arranged in a matrix with a pair on each row, and all words in the response remained on the screen until the eighth response word was finished.

Cued Recall

In cued recall, the first word in a pair appeared, and the participant was to type in the second word, according to the pairing that had been seen previously. In Experiment 2, because no pairing had been seen

TABLE 1
Experimental Conditions in Experiment 1

Condition	Training phase	List recall	Cued recall
No-study	Not included in this condition	One list (e.g., C-D, G-H, A-B, E-F)	Same pairs (e.g., G-??)
0-pairings	Word singletons A, B, C, D, E, F, G, and H four times each (random order)	One list (e.g., C-D, G-H, A-B, E-F)	Same pairs (e.g., G-??)
1-pairing	Word pairs A-B, C-D, E-F, and G-H once each and each word three times as a singleton (random order)	One list maintaining pairs (e.g., C-D, G-H, A-B, E-F)	Same pairs (e.g., G-??)
2-pairings	Word pairs A-B, C-D, E-F, and G-H twice each and each word twice as a singleton (random order)	One list maintaining pairs (e.g., C-D, G-H, A-B, E-F)	Same pairs (e.g., G-??)
4-pairings	Word pairs A-B, C-D, E-F, and G-H four times each (random order)	One list maintaining pairs (e.g., C-D, G-H, A-B, E-F)	Same pairs (e.g., G-??)

Note. For a given participant, different words appeared in each training condition. Presentations of words or pairs from one condition were usually separated by presentations of other words and pairs. Specific pairings, but not the order of pairs, were maintained throughout.

yet in the no-study and 0-pairings conditions, participants were instructed that they could simply respond "s" (to indicate they had seen the cue word before in the experiment) or "n" (to indicate they had not seen it). In these two conditions, a participant never happened to respond with the word that was subsequently paired with the cue in the list-recall phase.

RESULTS

In this section, we assess the effects of the manipulation of training repetition on list recall using two criteria: strict serial-position scoring, in which an item is counted correct only if it is recalled in the correct position, and free scoring, in which an item is counted correct if it is recalled anywhere in the list. Cued recall and its relation to list recall also are examined, to explore the nature of underlying associations. Then all of the data are examined to assess the constant-capacity hypothesis, first using raw data and then in a more exacting manner using a multinomial model. Last, the relation of chunk span to order information is examined.

Effects of the Training Manipulation

As shown in Figure 1 (top and middle panels), the manipulation of training condition was successful. For each experiment and scoring criterion, a 5 × 8 analysis of variance (ANOVA) of the list-recall scores including training condition and serial position (1–8) as within-subjects factors yielded a significant main effect of training condition. We estimate effect size with η_p^2, which is calculated as $SS_{\text{effect}}/(SS_{\text{effect}}+SS_{\text{error}})$. This statistic is independent of which other factors are included in the analysis. For Experiments 1 and 2, with serial-position scoring, the relevant statistics were as follows: $F(4, 60) = 10.51$ and 4.36, respectively; $\eta_p^2 = .41$ and .23. With free scoring, the statistics were $F(4, 60) = 16.55$ and 10.84, respectively; $\eta_p^2 = .52$ and .42.

As shown in Figure 2, the serial-position functions obtained (collapsed across training conditions) were typical of those obtained in serial-recall experiments (cf. Neath & Surprenant, 2002).

The manipulation of training condition also was effective in cued recall (Fig. 1, bottom panel). For Experiment 1, a 5 × 4 ANOVA was conducted with all five training conditions and four serial positions (corresponding to the serial positions of the pairs in list recall, which had been tested previously). This analysis produced a significant effect of training condition, $F(4, 60) = 19.10$, $\eta_p^2 = .56$, and of serial

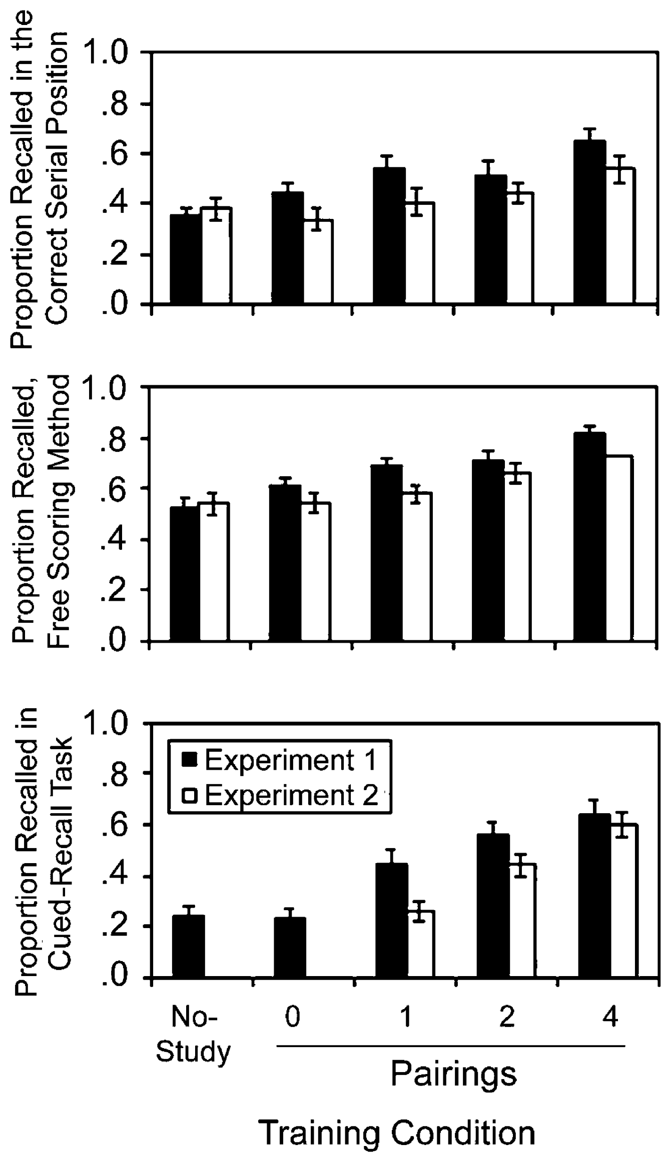

Fig. 1. Proportion of items correct in Experiments 1 and 2. The top panel shows results for list recall according to serial-position scoring, the middle panel shows results for list recall according to free scoring, and the bottom panel shows results for cued recall. Error bars denote standard errors.

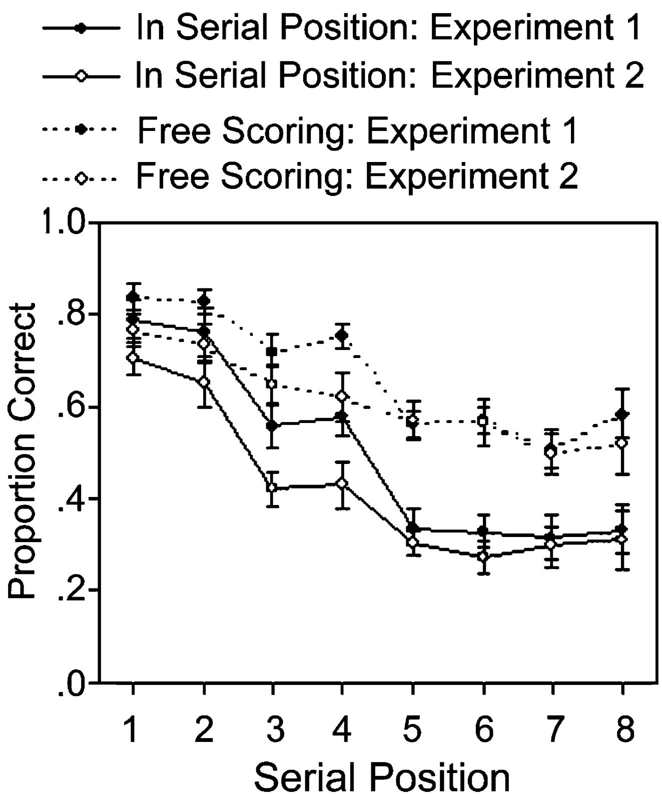

Fig. 2. Proportion of items correct in list recall as a function of serial position. Results are shown separately for serial-position scoring and free scoring (for which serial position in the response does not matter) in each experiment. Error bars denote standard errors.

position, $F(3, 45) = 6.70$, $\eta_p^2 = .31$. The interaction term was insignificant. The performance levels for the four serial positions (with standard errors) were .50 (.05), .47 (.04), .35 (.03), and .37 (.04). Thus, placement of items in early serial positions in list recall resulted in relatively high performance on those items not only in list recall, but also in cued recall when it occurred subsequently. For Experiment 2, the no-study and 0-pairings conditions were not included in the analysis because participants had no information about the pairings at the time of cued recall. A 3 (training condition) × 4 (serial position in subsequent list recall) ANOVA produced only a significant effect of training condition, $F(2, 30) = 28.36$, $\eta_p^2 = .65$. As expected, no serial-position effect occurred because, in this experiment, cued recall preceded any exposure to lists.

Next, we asked whether the learning demonstrated in cued recall was mirrored in list recall. To do so, we calculated additional information from list recall. We considered whether a pair was reproduced intact (i.e., with the two words adjacent and in the correct order, regardless of whether these words were recalled in their correct serial positions or shifted in the list) to be initial evidence that a pair was chunked in list recall. Figure 3 plots the mean proportion of intact pairs in list recall as a function of the mean proportion of correct cued recall, for every training condition in both experiments. If the two measures were based on exactly the same information, the data points would fall along the diagonal line. The fact that the points fall above the line indicates that there was some sort of memory that was usable in list recall but not in cued recall. This is a striking finding inasmuch as participants had to produce word pairs in list recall, but only the second item in a pair in cued recall. The more important difference between the procedures is apparently that only list recall was an immediate-recall procedure. Participants can assemble associations between items in a list without those associations necessarily remaining available for cued recall.

Assessment of the Constant-Capacity Hypothesis

If the numbers of pairs recalled intact within lists could be taken as indices of two-item chunks formed, then it would be possible to assess the constant-capacity hypothesis stating that although the number of two-item chunks increases as the frequency of paired presentations during training increases, the total number of chunks recalled (learned pairs plus singletons) stays constant. The bars in the top panel of Figure 4 show the mean number of intact pairs recalled (serial recall) per list as a function of training condition. It is clear that the number increased markedly across training conditions. In a one-way, within-subjects ANOVA of these scores in Experiment 1, the training condition variable was significant, $F(4, 60) = 20.17$, $\eta_p^2 = .57$. The strength of the monotonic increase between the 0- and 4-pairings conditions is obvious in the figure, and well supported by post hoc Newman-Keuls tests (which were significant for the no-study condition vs. the 1-, 2-, and 4-pairings conditions; for the 0-pairings condition vs. the 1-, 2-, and 4-pairings conditions; and for the 1- and 2-pairings conditions vs. the 4-pairings condition). In Experiment 2, the effect of training condition again was significant, $F(4, 60) = 13.46$, $\eta_p^2 = .47$, and the Newman-Keuls tests showed the same effects as in Experiment 1, except that the difference between the no-study and 1-pairing conditions was insignificant.

In each training condition, an estimate of the number of 1- plus 2-item chunks recalled (the chunk span) was obtained by subtracting the number of intact pairs recalled from the total number of items recalled. This is appropriate because each pair recalled intact

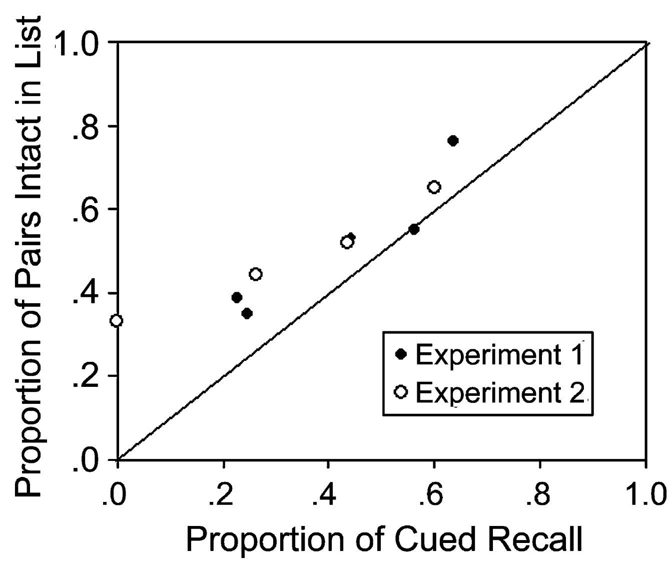

Fig. 3. Scatter plot of the mean proportion of intact pairs in list recall (either in the correct location or in an incorrect location in the list) as a function of the mean proportion correct in cued recall, for each condition in the two experiments.

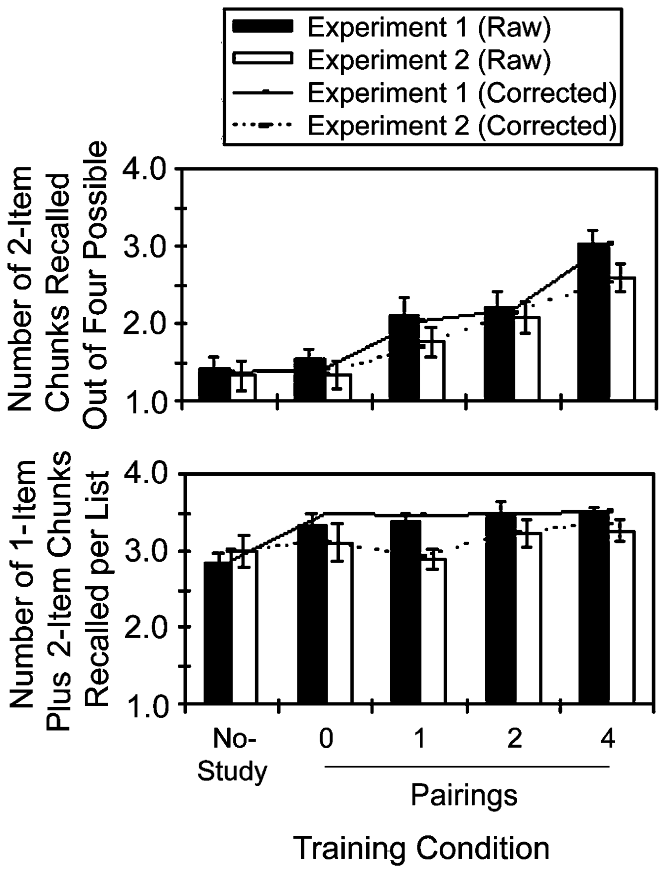

Fig. 4. Mean number of two-item chunks (top panel) and one- plus two-item chunks (bottom panel) recalled per list in the serial-recall task of each experiment. Scoring here does not take into consideration the serial position of the unit within recall output. Bars are based on raw data, and lines are based on data corrected using the multinomial model. Error bars denote standard errors for the raw data.

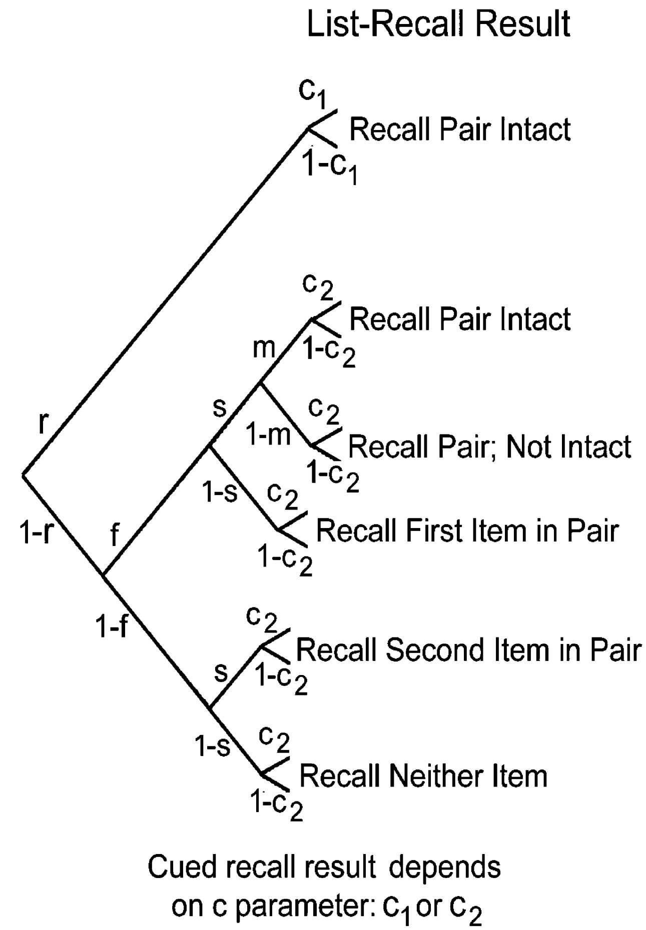

Fig. 5. Diagram of the multinomial model of performance in Experiment 1. The model for Experiment 2 was the same except that cued recall came before, rather than after, list recall. In the Experiment 2 model, there was only one c parameter, but there were different r and s parameters for the c and $(1 - c)$ branches.

accounts for the recall of two items. Thus, if 6.5 of 8 items were recalled on average and two of four 2-item pairs were recalled intact, then it would be estimated that 4.5 chunks (6.5 − 2), including 2.5 single-item chunks and the 2 two-item chunks, were recalled on average. The results of this analysis are shown by the bars in the bottom panel of Figure 4. Notice that the estimated number of chunks stays remarkably constant across the 0- through 4-pairings conditions despite the dramatic increase in 2-item chunks across conditions.

In the no-study condition, it is possible that individual words were difficult to retrieve inasmuch as confusion with semantic or phonological associates was possible. If a particular participant recalls some of the features of a word but not others, then, in effect, the word has not been encoded as a stable, single chunk. These factors would lower the estimate of capacity. With the no-study condition omitted, the differences between capacities in the other conditions did not approach significance in ANOVAs, either for the experiments combined or for each experiment separately. The slope of the estimates across the 0- through 4-pairings conditions for the two experiments combined was only 0.05 chunks per training presentation, with a 95% confidence interval of ± 0.06 chunks per presentation. The difference between the mean capacity in the 0- versus 4-pairings conditions was only 0.17 chunk (95% confidence interval of ± 0.30). Thus, any effect of paired-associate learning on the total number of chunks was quite small. Moreover, neither of these estimates was significantly different from zero. (In contrast, the number of two-item chunks recalled, which is not a capacity estimate, had a slope of 0.33 ± .07 chunks per presentation and a difference between the 0- and 4-pairings conditions of 1.39 ± 0.31 chunks.)

A problem with this analysis of capacity limits is that it is theoretically possible for a pair of words to be produced intact when, in fact, they have not been combined into a single chunk. It could happen that each of the two words is correctly recalled separately, resulting in a counterfeit chunk. To assess this possibility, we constructed ad hoc multinomial models (Batchelder & Riefer, 1999; Schweickert, 1993) of this task. Parameters of the models yielded corrected estimates of chunk span.

The model for Experiment 1 is depicted in Figure 5. Processes are represented as limbs and are assumed to occur in an all-or-none fashion, with the occurrence probabilities serving as free parameters. With probability r, the chunk is retrieved in list recall. If it is not retrieved then, with probability f, the first item in the pair is retrieved. Whether or not this takes place, with probability s, the second item in a pair is retrieved. If recalls of both the first and the second items occur independently, then with probability m, they are recalled in adjacent positions in the correct order, making them appear to form a chunk even though none actually has formed. In the subsequent cued-recall task, recall is successful with probability c_1 in the case in which chunk retrieval has been successful in list recall (with probability r), and with probability c_2 in all cases in which it has not been successful (with probability $1 - r$). The distinction between two c parameters can be justified on the grounds that prior successful chunk retrieval in list recall may strengthen the association.

The model for Experiment 2 (not shown) uses the same logic, but follows the temporal course of that experiment, in which cued recall preceded list recall. The model therefore begins with a single c parameter and then, in the case of either cued-recall success (c) or failure ($1-c$), follows with tree diagrams for list recall. Two list-recall parameters are conditioned on cued-recall success (r_1 and s_1 are used) or failure (r_2 and s_2 are used), whereas single parameters for m and f are used because they have no obvious dependence on cued-recall performance.

The models were fit to the list-recall and cued-recall data taken jointly, by the technique of maximum likelihood[1] (Riefer & Batchelder, 1988). The list-recall data designated intact recall of word pairs (regardless of whether they were recalled in the correct serial positions or were shifted in the list) but otherwise used free scoring.

The measurement model indicates that it was rare for an intact pair of items to be recalled when a chunk had not actually been formed. The probability of this happening is estimated by $[(1-r)fsm]$ in Experiment 1 and by $[c(1-r_1)fs_1m + (1-c)(1-r_2)fs_2m]$ in Experiment 2. The probability of a genuine two-item chunk being formed, estimated simply as r in Experiment 1 and $[cr_1+(1-c)r_2]$ in Experiment 2, was much higher. In fact, the percentage of intact pairs that could be attributed to true chunking according to this model was 98% overall, and 91% or higher in every condition of each experiment.

In the top panel of Figure 4, the corrected estimates of true two-item chunks (with counterfeit chunks excluded) are shown for each condition of the two experiments by the solid and dashed lines, respectively. The estimates are nearly identical to those obtained from the raw data, ruling out counterfeit chunks as a problem. Similarly, when the corrected estimates of two-item chunks were used to correct the estimates of the total number of chunks (Fig. 4, bottom panel, solid and dashed lines), again there was no substantial deviation from the uncorrected scores. Note that the fairly constant capacity shown in the bottom panel of Figure 4 fell out of the model rather than being built into it, lending support to the constant-capacity hypothesis. The particular magnitude of that capacity (between three and four chunks), and the need to use familiarized items to observe that capacity, are all consistent with the regularities pointed out by Cowan (2001).

Item and Order Information

Paired associations typically led not only to better recall of items in lists, but also to excellent serial-position accuracy. One can estimate the proportion of items recalled out of place by subtracting the proportion correct in serial-order scoring from that in free scoring (means in Fig. 1). The differences for the no-study and the 0-, 1-, 2-, and 4-pairings conditions were .17, .16, .15, .20, and .17, respectively, in Experiment 1. In Experiment 2, they were .16, .21, .17, .22, and .19. Notice the absence of increases in the differences across training conditions. This implies that effects of paired-associate learning included only increases in the number of correctly placed items, not increases in erroneously placed items. We therefore suggest that learned chunks are by nature context-specific (i.e., bound to the correct serial positions).

[1]We minimized the negative log-likelihood function with the simplex routine (Nelder & Mead, 1965). Over both experiments, the data representation had 80 degrees of freedom, with 59 model parameters. The log-likelihood ratio test statistic (G^2, $df=21$) of 39.2 slightly exceeds the .05 criterion of 32.7, indicating a slight misfit. However, the RMSE (root mean squared error) difference between predicted and obtained proportions of specific outcomes was only .010. Thus, the predictions are precise enough to estimate counterfeit chunks.

DISCUSSION

We have addressed a previously neglected but fundamental hypothesis based on Miller (1956). In this study, although inducing associations between words increased the number of two-word chunks, the total chunk span (the number of singletons plus two-word chunks) remained fairly constant. Both this constancy and the observed chunk span closely match expectations of Cowan (2001). Auxiliary findings are that learned chunks appear to include serial-position information, and that some chunks can be available for list recall without being strong enough to allow correct responding in subsequent cued recall.

The constant-capacity hypothesis is a bold hypothesis, and it remains to be seen if it will hold up across all other test circumstances. Discovering domains in which the constant-capacity hypothesis holds versus domains in which it does not hold will be theoretically valuable. It may well fail when attention must be shared between storage and processing (e.g., Daneman & Carpenter, 1980). There is a need for further theorizing in this field, and it is not yet clear what the capacity-limited holding mechanism is. Two possibilities are the focus of attention (Cowan, 2001) and an episodic buffer (Baddeley, 2001). Ultimately, a more explicit theoretical model of capacity limits is needed (for suggestions, see Luck & Vogel, 1998; Usher, Haarmann, Cohen, & Horn, 2001). We hope that these findings inspire further research, long overdue, examining the nature of capacity limits in immediate-recall tasks.

Acknowledgments—This work was supported by National Institutes of Health Grant R01 HD-21338 to N. Cowan and National Science Foundation Grant SES–0095919 to J. Rouder, D. Sun, and P. Speckman.

REFERENCES

Anderson, J.R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341–380.

Baddeley, A. (2001). The magic number and the episodic buffer. *Behavioral and Brain Sciences*, *24*, 117–118.

Baddeley, A.D. (1986). *Working memory* (Oxford Psychology Series No. 11). Oxford, England: Clarendon Press.

Batchelder, W.H., & Riefer, D.M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86.

Bjork, E.L., & Healy, A.F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*, 80–97.

Bowles, A.R., & Healy, A.F. (2003). The effects of grouping on the learning and long-term retention of spatial and temporal information. *Journal of Memory and Language*, *48*, 92–102.

Broadbent, D.E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18). London: Wiley.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cunningham, T.F., Marmie, W.R., & Healy, A.F. (1998). The role of item distinctiveness in short-term recall of order information. *Memory & Cognition*, *26*, 463–476.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Detterman, D.K. (1977). A comparison of item and position probes in short-term memory. *American Journal of Psychology*, *90*, 45–53.

Ericsson, K.A., Chase, W.G., & Faloon, S. (1980). Acquisition of a memory skill. *Science*, *208*, 1181–1182.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*, 200–209.

Frick, R.W. (1989). Explanations of grouping effects in immediate ordered recall. *Memory & Cognition*, *17*, 551–562.

Gobet, F., Lane, P.C.R., Croker, S., Cheng, P.C.-H., Jones, G., Oliver, I., & Pine, J.M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*, 236–243.

Healy, A.F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 644–655.

Hitch, G.J., Burgess, N., Towse, J.N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology*, *49A*, 116–139.

Hockey, R. (1973). Rate of presentation in running memory and direct manipulation of input-processing strategies. *Quarterly Journal of Experimental Psychology (A)*, *25*, 104–111.

Lee, C.L., & Estes, W.K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 149–169.

Lewandowsky, S., & Murdock, B.B., Jr. (1989). Memory for serial order. *Psychological Review*, *96*, 25–57.

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Luck, S.J., & Vogel, E.K. (1998). Response from Luck and Vogel. *Trends in Cognitive Sciences*, *2*, 78–80.

Mandler, G. (1985). *Cognitive psychology: An essay in cognitive science*. Hillsdale, NJ: Erlbaum.

Marmurek, H.H., & Johnson, N.F. (1978). Hierarchical organization as a determinant of sequential learning. *Memory & Cognition*, *6*, 240–245.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Murdock, B.B., Jr. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, *105*, 191–216.

Nairne, J.S., Whiteman, H.L., & Kelly, M.R. (1999). Short-term forgetting of order under conditions of reduced interference. *Quarterly Journal of Experimental Psychology (A)*, *52*, 241–251.

Neath, I., & Surprenant, A. (2002). *Human memory* (2nd ed.). Belmont, CA: Wadsworth.

Nelder, J.A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313.

Ng, H., & Maybery, M.T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology*, *55A*, 391–424.

Pollack, I., Johnson, I.B., & Knaff, P.R. (1959). Running memory span. *Journal of Experimental Psychology*, *57*, 137–146.

Riefer, D.M., & Batchelder, W.H. (1988). Multinomial modeling and the measure of cognitive processes. *Psychological Review*, *95*, 318–339.

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of groups. *Quarterly Journal of Experimental Psychology*, *21*, 137–147.

Saint-Aubin, J., & Poirier, M. (1999). The influence of long-term memory factors on immediate recall: An item and order analysis. *International Journal of Psychology*, *34*, 347–352.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 168–175.

Slak, S. (1970). Phonemic recoding of digital information. *Journal of Experimental Psychology*, *86*, 398–406.

Towse, J.N., Hitch, G.J., & Skeates, S. (1999). Developmental sensitivity to temporal grouping effects in short-term memory. *International Journal of Behavioral Development*, *23*, 391–411.

Tulving, E., & Patkau, J.E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology*, *16*, 83–95.

Usher, M., Haarmann, H., Cohen, J.D., & Horn, D. (2001). Neural mechanism for the magical number 4: Competitive interactions and non-linear oscillations. *Behavioral and Brain Sciences*, *24*, 151–152.

Wickelgren, W.A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, *19*, 97–102.

Wilson, M. (1987). *MRC psycholinguistic database: Machine usable dictionary. Version 2.00*. Chilton, Didcot, England: Rutherford Appleton Laboratory, Science and Engineering Research Council, Informatics Division.

(Received 7/10/03; Revision accepted 9/2/03)