THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1996, *49A* (1), 80–115

Unchained Memory: Error Patterns Rule out Chaining Models of Immediate Serial Recall

Richard N. A. Henson, Dennis G. Norris,
Michael P. A. Page, and Alan D. Baddeley
MRC Applied Psychology Unit, Cambridge, U.K.

Many models of serial recall assume a chaining mechanism whereby each item associatively evokes the next in sequence. Chaining predicts that, when sequences comprise alternating confusable and non-confusable items, confusable items should increase the probability of errors in recall of following non-confusable items. Two experiments using visual presentation and one using vocalized presentation test this prediction and demonstrate that: (1) more errors occur in recall of confusable than alternated non-confusable items, revealing a "sawtooth" in serial position curves; (2) the presence of confusable items often has no influence on recall of the non-confusable items; and (3) the confusability of items does not affect the type of errors that follow them. These results are inconsistent with the chaining hypothesis. Further analysis of errors shows that most transpositions occur over short distances (the locality constraint), confusable items tend to interchange (the similarity constraint), and repeated responses are rare and far apart (the repetition constraint). The complete pattern of errors presents problems for most current models of serial recall, whether or not they employ chaining. An alternative model is described that is consistent with these constraints and that simulates the detailed pattern of errors observed.

How is a sequence of items, such as a telephone number, stored in memory and recalled in the correct order? One class of theories assumes that learning a sequence involves the formation or strengthening of associations between representations of successive items (e.g. Ebbinghaus, 1964; Wickelgren, 1965). Recall can proceed by stepping through these associations in a process called chaining. Given a sequence *A*, *B*, the simplest form of chaining involves using the response of *A* as a cue for the retrieval of its associate *B*. Chaining of some form has remained a popular means of ordering recall from memory (e.g. Jordan, 1986; Lewandowsky & Murdock, 1989; Richman & Simon, 1994).

Requests for reprints should be sent to Richard Henson, MRC Applied Psychology Unit, 15, Chaucer Road, Cambridge, CB2 2EF, U.K. Email: rik.henson@mrc-apu.cam.ac.uk

The first author was supported by a MRC Research Studentship while registered for a PhD at the University of Cambridge. Part of the work was supported by a Joint Councils Initiative Grant to Dennis Norris, Alan Baddeley, and Graeme Mitchison, Grant Number E304/187. We would like to thank Peter Watson and Ian Nimmo-Smith for statistical advice.

Several immediate objections to simple chaining models arise. For example, how are sequences with repeats recalled, where two items share the same cue? Or how do people recover from errors, where the functional cue for the response following an error will differ from the cue for a correct response—"a chain is only as strong as its weakest link"? However, more sophisticated chaining models can overcome these problems. Murdock's TODAM model (Lewandowsky & Murdock, 1989; Murdock, 1987) allows recovery from errors by effectively cuing with the previous response only if it is correct. Otherwise, a cue approximating the correct one is used. Sequences with repeated items are less problematic when models assume associations over more than just adjacent items, so that the cue beomes a compound of a number of previous responses. Then the effect of a single error is also less devastating. This type of compound cuing has become very popular in recurrent neural network models (e.g. Elman, 1990; Jordan, 1986; Taylor, 1991). Another alternative is to appeal to the type/token distinction, so that two occurrences of the same type in a sequence have non-identical token representations. For example, item representations may be embedded in different temporal or spatial contexts, allowing the same item to function as a different cue at different positions in a sequence, as in the case of Wickelgren's "allophones" (Wickelgren, 1969).

Nevertheless, many researchers have argued strongly against the sufficiency of any type of chaining as a general account of sequential behaviour (e.g. Jensen & Rohwer, 1965; Johnson, 1972; Lashley, 1951). For example, how can a person's memory for the spelling of thousands of English words be stored only as set of associations between 26 letters? If this were the case, the degeneracy of these associations would surely predict massive interference in the action of writing. Many theories rely instead on positional cues (e.g. Conrad, 1965; Shiffrin & Cook, 1978; Slamecka, 1967) or cyclic reactivations (e.g. Estes, 1972; Lee & Estes, 1977, 1981). Burgess and Hitch's (1992) neural network model allowed the modellers to vary the relative influence of chaining versus positional cuing. Their best fits to data from a number of short-term memory experiments were obtained when chaining was effectively absent. However, although chaining may be insufficient and even problematic in some formulations, the question addressed here is whether there exists any unequivocal, empirical evidence for chaining in immediate serial recall.

Related attempts to find evidence for associations between items in a learned sequence have looked for transfer effects between serial and paired-associate learning. Unfortunately, conclusions have been mixed (Young, 1968); moreover, the appropriateness of applying results from this paradigm to immediate serial recall remains unclear. A far simpler way to find evidence for chaining is to examine the nature of errors that people make when attempting serial recall. As Estes (1972) observed: "When retention is imperfect, the confusion errors that occur are highly systematic" (p. 161). Although a single error may reflect a temporary failure to realize an accurate representation in memory, regularities and patterns in the distribution of large numbers of errors shed light on the mechanisms subserving recall and consequently the memory representations on which these mechanisms operate.

An example of such error analysis is Wickelgren's (1966) demonstration of "associative intrusions". These errors reflect a greater proportion of transpositions between items in positions following repeated items than in identical positions in control lists without repeats. Wickelgren attributed this difference to chaining along forward associations between representations of adjacent items: Repeated items are ambiguous cues because they possess more than one forward association. The problem with this measure is that, although errors are only counted as associative intrusions when the preceding repeat has been recalled correctly, the presence of repeats may affect the baseline chance of many possible transpositions, not just those following repeats. Considerable evidence suggests that the encoding of repeated items may differ to that of other items in memory (e.g. Jahnke, 1969; Lee, 1976). Repeats are often noted explicitly by people, in which case they may lead to different grouping strategies. For example, if repeats signalled the start of a new subjective or rehearsal group, associative intrusions could be the result of systematic transpositions between groups (e.g. Ryan, 1969). Nevertheless, if ambiguous cues do exert a real effect, that effect may be demonstrable with items that, though not repeated, are phonologically similar.

Phonological Similarity

An abundance of empirical data suggests that representations underlying performance in most verbal short-term memory tasks are speech-based. The order of items that are pronounced similarly (even if they are read in silence), such as *B*, *D*, *G*, *P*, is more difficult to recall than is the order of items that are pronounced differently, such as *C*, *F*, J, *R* (e.g. Baddeley & Ecob, 1970; Conrad & Hull, 1964). This *phonological similarity effect* (Baddeley, 1986) occurs in spite of the fact that the items themselves are often more likely to be recalled when similar, albeit in the wrong order, as can be demonstrated by comparing serial with free recall (Watkins, Watkins, & Crowder, 1974). One reason for order errors could be that phonologically similar items are likely to be confused when selecting a response for a given position (i.e. an effect of similarity at retrieval). In addition, models that chain along phonological representations predict that any similarity between cues for different items also causes confusion (i.e. an effect of similarity at cuing). Wickelgren (1965) for example, attributed the phonological similarity effect to similar items sharing one or more phonemes, each phoneme with a single representation in memory. If chaining proceeds along such representations, the situation becomes formally equivalent to a sequence with repeated items, and the phonological similarity effect arises for the same reason as Wickelgren's associative intrusions. Thus, in sequences like *B*, J, *D*, *R*, the correct cue for J (*B*), is phonologically similar to the cue for *R* (*D*), which would lead to uncertainty for the response that should follow *B*.

Baddeley's (1968) Experiment V was an attempt to distinguish effects of similarity at retrieval from effects of similarity at cuing. He tested immediate serial recall of lists of six visually presented consonants. The lists were constructed from an experimental vocabulary consisting of a set of consonants that were all pronounced similarly (the confusable items) and a set of consonants pronounced differently (the non-confusable items). With lists in which confusable and non-confusable items alternated, Baddeley found that more errors occurred in recalling confusable items than in recalling non-confusable items. The differential error rate in such *alternating lists* was revealed by the "sawtooth" appearance of error position curves (graphs of error percentages per serial position—see, for example, Figure 1), in which the peaks of the sawteeth represented errors in recall of confusable items and the troughs represented fewer errors in recall of non-confusable items.

Moreover, the sawteeth were confined within more conventionally bowed curves for two types of *pure lists*: the *confusable lists*, which contained only confusable items, and the *non-confusable lists*, which contained only non-confusable items. The peaks of the sawteeth lay below the error position curve for confusable lists, but the troughs were virtually coincident with the curve for non-confusable lists.

Baddeley took the fact that most errors in recall of alternating lists occurred for confusable rather than non-confusable items, as favouring the idea that phonological similarity acts at retrieval rather than at cuing. Indeed, the presence of confusable items in alternating lists seemed to have no effect on the probability of recalling the non-confusable items when compared with lists of all non-confusable items, suggesting that no effect of phonological similarity at cuing exists at all.

However, disregarding chaining models on the basis of these results is premature for a number of reasons. The sawteeth on their own are certainly insufficient. This is because chaining models such as TODAM could predict an effect of similarity at retrieval as well as at cuing. Sawteeth could then result if the effect of phonological similarity is simply much greater at retrieval than at cuing. The apparent coincidence of alternating and non-confusable curves, for recall of non-confusable items, is much more difficult to reconcile with chaining models. However, this coincidence was not found in Experiment VI of the same paper, which used positional recall of auditorily presented words. Nor was it found on all positions for Baddeley's lists of three confusable followed by three non-confusable items, in which the percentage errors for the first non-confusable item appeared slightly greater than for the corresponding non-confusable item in non-confusable lists. More importantly, even exact coincidence of alternating and non-confusable curves is inadequate to rule out an effect of phonological similarity at cuing. This is because error position curves fail to distinguish between a cue containing correct previous responses and a cue that contains erroneous previous responses.

Consider an erroneous response in recall of an alternating list, where the correct confusable item is replaced by another confusable item (perhaps through phonological similarity acting at retrieval). According to chaining models, this response forms part (or all) of the cue for the next response. However, because the confusable item makes the functional cue similar to cues for other possible responses, there is a chance that the correct non-confusable item will follow in spite of this error. On the other hand, if the last response were an erroneous non-confusable item instead, the chance of recovering from this error would be less. Rather, the non-confusable item would more probably be followed by another erroneous response—specifically, its successor in the list. In other words, the next response would tend to be in the correct relative order to the last response, though both would be errors in absolute order. Thus an effect of similarity at cuing may not only increase the chance of a first error, but also the chance of recovering from an error. Given that the chance to recover from an error in recall of non-confusable lists is less than in alternating lists, similar error percentages for non-confusable items in the different list types are possible. This possibility is shown more formally in the Appendix.

A stronger test of models that chain along phonological representations is to restrict analysis to only those cases where the functional cue for the next response is identical to the correct cue for the next response. This is the case when all previous responses are correct. Calculating the proportion of errors on each position that are the first errors to occur in reports allows estimation of the conditional probability of failing to recall an item, given that previous responses were correct. The percentage of such *first-in-report* errors can be plotted as conditional error position curves. If conditional error proportions for non-confusable items are still the same following correct recall of a confusable item in alternating lists as following correct recall of non-confusable item in non-confusable lists, then there would be no evidence for an effect of phonological similarity at cuing (given present experiments meet Frick's 1995 criteria for accepting this null hypothesis). This would be incompatible with models that chain along phonological representations. On the other hand, a significant difference will lend support to chaining theories. Testing this hypothesis was the first aim of the present experiments.

Error Types

A second aim was to conduct a more thorough analysis of subjects' responses than is conventionally attempted. Though Baddeley reported errors by serial position, he did not examine actual types of error: for example, whether the errors were *omissions* or *substitutions*. Substitutions could also have been either *intrusions* of items not appearing in the current list (often intruding from previous lists) or *transpositions*, which are reorderings of list items. Such analysis of errors addresses further theoretical questions. For example, some theories suggest that similar representations degrade faster than do dissimilar ones, as in Posner and Konick's (1966) "acid bath" theory. In this case, the peaks of the sawteeth in Baddeley's data may simply have reflected a greater incidence of confusable items being omitted, or being substituted by random guesses from the experimental vocabulary. However, if phonological similarity acts through response competition during retrieval (e.g. Baddeley, 1968, Experiments I–IV) and perhaps rehearsal (e.g. Murdock & vom Saal, 1967), then the majority of these errors should reflect one confusable item substituting for another confusable item (e.g. Bjork & Healy, 1974; Conrad, 1965). As shown in the Appendix, this type of substitution is important if chaining theories are to be reconciled with Baddeley's data.

Selective substitution of confusable items implies that the peaks of the sawteeth should remain even if the only errors were transpositions. In all three experiments described here, lists were generated by a small experimental vocabulary, conforming to the "order only" condition of Healy (1974). With such a design, subjects know in advance which particular items they will see in a given trial and so need only concentrate on the order in which they occur. Consequently, the number of intrusions and omissions should be minimal, and analysis can be focused on the problem of retaining the order of items. Making the simplifying assumption that subjects' reports are permutations of list items also allows determination of some chance probabilities of correct ordering of items.

Further classification of transpositions is possible. Transposition matrices can be constructed to show the proportion of responses in which an item from a given position in a list (its *input* position) is produced at a given position in a report (its *output* position). Correct responses appear on the leading diagonal of this matrix, forming the conventional serial position curve; all other entries represent transposition errors. Many researchers (e.g. Lee & Estes, 1977) have reported a tendency for transpositions to be localized around the correct position, as revealed by a monotonically decreasing gradient of proportions of transpositions against transposition distance (the difference between input and output positions).

An important subclass of transpositions consists of erroneous repeated responses (when no items were repeated in stimulus lists). If such *repeat errors* are rare, then responses are probably being made from the experimental vocabulary virtually without replacement. As a consequence, dependencies will exist between errors made across output positions, which will be reflected in different shapes of unconditional (conventional) versus conditional error position curves. Such dependencies have important implications for analysis of errors. A second subclass of transpositions consists of *relative* errors. A relative error is an erroneous response in the correct order relative to the previous response (which must therefore also be an error). A second test of chaining models, given the above examples of the effect phonological similarity may have at cuing, is to compare the frequency of relative errors in different list types. Models chaining along phonological representations predict that relative errors will be more frequent for non-confusable than confusable lists, because an erroneous non-confusable item is more likely to cue its successor in the list than is an erroneous confusable item.

In summary, although the serial position curve is highly informative, a great deal more information is made available by studying the detailed pattern of errors. Whether or not such information supports chaining models, the patterns of errors made by subjects should further constrain possible models of serial recall.

EXPERIMENT 1

The first experiment was a near-replication of four conditions in Baddeley's (1968) Experiment V: the confusable lists, the non-confusable lists, and two types of alternating list, depending on whether alternation begins with a confusable or a non-confusable item. One important difference from Baddeley's experiment was that intrusion and omission errors were minimized by blocking together trials on the different types of list rather than randomly interspersing them, and ensuring that all lists in a block contain the same six items (simply arranged in a different order for each trial). A second difference in design was that subjects in the present experiment were encouraged to group the six items into two groups of three. Baddeley did not report giving such instruction to his subjects. However, grouping strategies are often brought to bear on the simplest of span tasks (Frankish, 1974), and they can have important effects on the pattern of transpositions. Particular advantage is conveyed to recall of the first and last items in a group, sometimes revealed as mini-primacy and mini-recency effects within groups. In fact, such a suggestion of spontaneous grouping by subjects is apparent in Baddeley's error position curves, particularly across Positions 3 and 4 of confusable curves. Different grouping strategies may interact differently with the structure of alternating lists. For example, a choice of grouping in twos rather than threes may have an effect on the nature of errors made in recalling alternating lists. Thus the explicit instruction to group in threes in the present experiments was intended to encourage a single, consistent grouping strategy across subjects.

The aims of the experiment were: (a) to reproduce and make explicit tests of Baddeley's findings, specifically the sawtooth error position curves for alternating lists; and (b) conduct a more thorough analysis of patterns of transpositions, repeat errors, and relative errors.

Method

Subjects

Forty-eight subjects from the APU Subject Panel were tested, of whom 17 were male and 31 were female. Their mean age was 27 years.

Materials

Stimuli were lists of 6 single-syllable consonants. They were generated from an experimental vocabulary of 12 consonants, which were classified according to their confusability—that is, whether they were phonologically similar to any other consonants in the vocabulary. The 6 confusable consonants shared a common rhyme when pronounced, *B, D, G, P, T, V*, and the 6 non-confusable consonants possessed unique rhymes, *H, K, M, Q, R, Y*.

The two pure list types were the confusable lists (*PC*), containing all 6 confusable consonants, and the non-confusable lists (*PN*), containing all 6 non-confusable consonants. Two alternating list types (*A1* and *A2*) were identified according to the two mutually exclusive sets of three confusable and three non-confusable consonants in the vocabulary. Alternating lists were further classified according to whether the alternation began with a confusable or a non-confusable item in the first position (*AC* and *AN* list types, respectively; see Table 1). List types *AC* and *AN* were nested inside list types *A1* and *A2*, such that a block of *A1* or *A2* lists contained 6 lists of type *AC* and 6 of type *AN*. With the randomized order of lists within blocks, this nesting was to reduce the chance of subjects' detecting a pattern of confusable/non-confusable alternation (e.g. as might occur if lists in a block always started with a confusable item).

The lists themselves were generated according to the following constraints: None of the lists contained obvious acronyms (nor co-occurrence of letters in alphabetical order), each consonant appeared equally often (twice) in each position, and the frequency of adjacent consonant pairs across the lists was made as uniform as possible, after the above considerations had been met. In other words, first- and second-order contingencies over the lists were close to being balanced.

TABLE 1
Composition of List Types in Experiment 1

List Type	*List Structure*[a]	*Letter Set (Example List)*	*No. of Lists*
PC	CCCCCC	BDGPTV	12
PN	NNNNNN	HKMQRY	12
A1, AC	CNCNCN	DQTMPK	6
A1, AN	NCNCNC	QDMTKP	6
A2, AC	CNCNCN	BHGYVR	6
A2, AN	NCNCNC	HBYGRV	6

[a] C = confusable item, N = non-confusable item.

Procedure

Every subject attempted recall of 4 blocks of 12 lists, each block containing lists of one of the list types *PC*, *PN*, *A1*, or *A2*. Before the first list of each block, the 6 consonants that would be seen in the following 12 trials were presented in a circle, in order to familiarize subjects with the set of possible responses. Subjects were told that the lists contained no repeated letters. The trial order of the 12 lists within blocks was randomized, and the block order was fully counter-balanced across subjects.

The experiment was run on an IBM PC, with the capitalized letters appearing in the centre of a monochrome VDU, each letter about half an inch high and replacing the previous one. Presentation rate was 400 msec per item, with a 100-msec interstimulus interval. Subjects were instructed to read the letters in silence. Immediate written recall was allowed after the last item disappeared, prompted with the display *Please Recall the List Now*! Subjects wrote one letter in each box of a row of six provided on a response sheet. A minimum of 10 sec was required between trials, after which subjects had to press a key to start the next trial. A short break of about a minute occurred between blocks.

Subjects were instructed to write down answers immediately and, if unsure, told simply to "write the first letter that comes to mind". If they really could not recall a letter, they were asked to put a line through the appropriate box. Subjects were reminded to recall in a forward manner, writing from left to right on the response sheet, and to resist the temptation to recall the last few letters first. Finally, subjects were advised that grouping the six letters into two groups of three may aid their retention; an example of such *3–3* grouping of a telephone number was given. Three practice trials then followed. The whole experiment took about 20 minutes.

Results and Discussion

The error position curves for Experiment 1 are shown in Figure 1. The present results clearly replicated the main features of Baddeley's data. There was a strong effect of phonological similarity, with confusable lists being more difficult to recall than non-confusable lists, and the alternating lists produced the same sawtooth pattern observed by Baddeley. The most important feature of the data is that there was no evidence that the non-confusable items in alternating lists were harder to recall than when they appeared in pure non-confusable lists. In fact, rather surprisingly, non-confusable items in alternating lists were actually recalled slightly better than those in pure non-confusable lists (i.e. the sawteeth straddled the non-confusable curve, rather than sitting on top of it). Closer inspection of the stimuli suggested a reason for this: The consonants in different list types differed in their *predictability*.

Baddeley, Conrad, and Hull (1965) have shown that predictable sequences of letters are easier to remember than are less predictable sequences. Their measure of predictability was derived from subjects' guesses for successive letters. An alternative, objective index of predictability is the frequency of occurrence of letter bigrams in written English (e.g. Baddeley, 1971). Lists containing common bigrams are likely to be easier to remember than those containing uncommon bigrams. Taking the logarithms of the number of occurrences of each of the five consonant bigrams per list in a corpus of over one million words collected by Solso and Juel (1980), the predictability of the consonants in Input Positions 2–6 was calculated (see Table 2). A two-way analysis of variance (ANOVA) on these predictabilities showed a significant main effect of list type, $F(3, 220) = 26.8$, $p < 0.001$, with consonants in *A2* alternating lists being most predictable on average, but

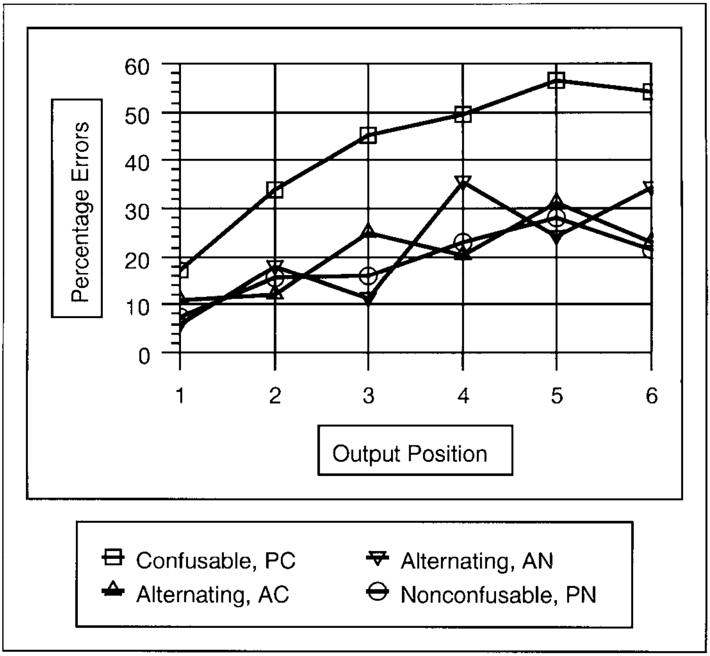

FIG. 1. Error position curves for the four list types in Experiment 1.

no effect of input position, $F(4, 220) = 0.1$, $p = 0.98$, or interaction with list type, $F(12,220) = 0.1$, $p > 0.99$. The fact that consonants in *A2* lists were much more predictable than those in other lists may explain why there were fewer errors in blocks of these lists (*mean* = 11.8, *SD* = 10.9) than either *A1* lists (*mean* = 18.3, *SD* = 12.9) or even non-confusable lists (*mean* = 13.4, *SD* = 12.2). The effect of relatively predictable *A2* lists will then be to reduce the level of the alternating curves relative to the non-confusable curve.

Because of the confounding of predictability, one cannot attach too much significance to the relative levels of performance on difference list types. Indeed, although the larger experimental vocabulary in Baddeley's experiment makes it less likely that there was a confound between predictability and list type in his experiment, a similar caution should apply to his results also. Without knowing whether predicability was equated across conditions in Baddeley's experiment, one cannot be entirely confident that performance on non-confusable items in alternating lists really was unaffected by the presence of confusable items. The issue of whether performance on non-confusable items is influ-

TABLE 2
Consonant Predictability for List Types in Experiment 1

	PC	*PN*	*A1*	*A2*
mean	1.48	1.73	1.08	2.73
SD	0.93	1.43	0.78	0.79

enced by the presence of the confusable items was addressed in Experiments 2 and 3, where predictability was controlled. However, the lack of a main effect of input position on predictability or interaction with list type means that predictability should not affect tests across positions within lists. The following within-list analyses are thus not undermined by between-list differences in predictability.

Overall Performance

Approximately 58% of non-confusable lists and 20% of confusable lists were recalled correctly. However, there was considerable variation in subjects' performance. For example, the number of errors made per non-confusable list varied from 0 to 3.5 (*mean* = 1.1, *SD* = 1.0). Of the 3566 errors in total, omissions represented approximately 5% of the errors, and intrusions amounted to only 3%.

Error Position Curves

Errors were broken down by output position. In order to test the reproducibility of Baddeley's findings, a number of planned comparisons were performed across output positions for each list in Figure 1, in four separate ANOVAs on log-odds scores (using an empirical logistic transform that caters for floor and ceiling effects in proportions and allows weighting of proportions by their sample size; see Cox & Snell, 1989).

The linear, orthogonal comparisons for the pure lists, *PC* and *PN*, tested for primacy (the average error score on Positions 1 and 2 compared with the average on Positions 3 and 4) and the last-item recency that is characteristic of serial recall of visual items (the average error score on Position 6 versus Position 5). Both non-confusable and confusable curves showed significant primacy effects, $F(1, 235) = 81.1$, $p < 0.001$, and $F(1, 235) = 18.0$, $p < 0.001$, respectively, but only the non-confusable curve showed a significant recency effect, $F(1, 235) = 8.2$, $p < 0.01$.

Three contrasts for alternating lists *AC* and *AN* tested the hypothesis that more errors occurred on confusable positions than on adjacent non-confusable positions. A fourth contrast looked for an effect of primacy over the first four positions. For both alternating curves, significantly greater numbers of transpositions were made on confusable positions than on adjacent non-confusable positions, $F(1, 195) > 4.6$, $p < 0.05$ in all cases, except between the first two positions of *AC*, $F(1, 235) < 0.1$, $p > 0.99$. Significant primacy effects occurred in both *AC*, $F(1, 235) = 43.3$, $p < 0.001$, and *AN*, $F(1, 235) = 42.3$, $p < 0.001$.

In summary, the error position curves are very similar to those found by Baddeley in 1968, apart from a lower overall rate and the influence of predictability. The lower error rate probably comes from a much lower incidence of intrusions and omissions than in Baddeley's experiment, given the smaller experimental vocabulary. Lower error rates (rarely more than 50% per position) are probably desirable, in that they are more likely to reflect systematic errors rather than purely random guesses. Moreover, the primacy, recency, and sawtooth effects apparent in both studies have been tested explicitly in the present experiment and found significant in nearly all cases, apart from the lack of significant recency in confusable curves, suggesting an interaction between phonological similarity and recency (which has been suggested elsewhere, e.g. Drewnowski, 1980). The only other case where a test was not significant was when effects of phonological similarity and primacy were in opposition (e.g. for the first two positions of *AC*).

Prima facie, these results generally support Baddeley's conclusions of a large effect of phonological similarity at retrieval, but only a small effect, if any, at cuing. However, a more conclusive test of phonological similarity at cuing must await the conditional analysis performed in Experiment 2, where predictability of different list types was balanced.

Transposition Matrices

Errors were further broken down by input position. The aggregate transposition matrices for *PC*, *PN*, and *AN* list-types, totalled across all subjects, are displayed graphically in Figure 2. In each panel, the six vertical bars for each output position represent the proportions of responses in that output position that came from Input Positions 1–6 (from left to right). These bars form the six transposition gradients for each output position. For the confusable and non-confusable lists (upper and middle panels), transposition gradients show a basic triangular shape, whereby the proportion of total responses transposed from a given input position fell off monotonically with distance between that input position and the output position. This monotonic decrease was remarkably lawful. For example, the descending rank ordering of proportions, seen most clearly from Input Positions 2–6 to Output Position 1, might be expected to occur only 1 in 120 times, if subjects forgetting the correct item guessed randomly from the set of list items. It is more likely that the pattern reflects some imperfect storage or retrieval of order information. The only exceptions to this monotonic decrease occurred for transpositions from early input positions to later output positions in recall of non-confusable lists (e.g. from Input Position 1 to Output Position 6 in the middle panel). However, further inspection showed that these exceptions were due to the increasing number of repeat errors from early input positions (see next section). Interestingly, the transposition gradients for non-confusable and confusable lists, for Output Position 1 for example, were not parallel: The gradients were steeper for confusable curves.

The transposition gradients for the *AN* alternating lists (lower panel) were not always monotonically decreasing, but, rather, proportions depended on the phonological similarity between the correct and transposed item. Thus output positions that corresponded to confusable input positions (Positions 2, 4, and 6 here) showed small peaks for other confusable input positions. The same pattern arose for *AC* lists. Because the majority of reports were in effect permutations of list items, given that most errors were non-repeated transpositions, the significantly greater proportion of errors on confusable than on non-confusable positions in error position curves implies that the majority of transpositions made in recall of alternating lists were confusable items transposing with other confusable items. This supports the notion of similarity acting at retrieval, specifically through active response competition, rather than passive decay or interference in theories like the acid bath theory (Posner & Konick, 1966).

In summary, transposition gradients show sensitivity to both transposition distance and phonological similarity. The pattern of transpositions can be summarized by two

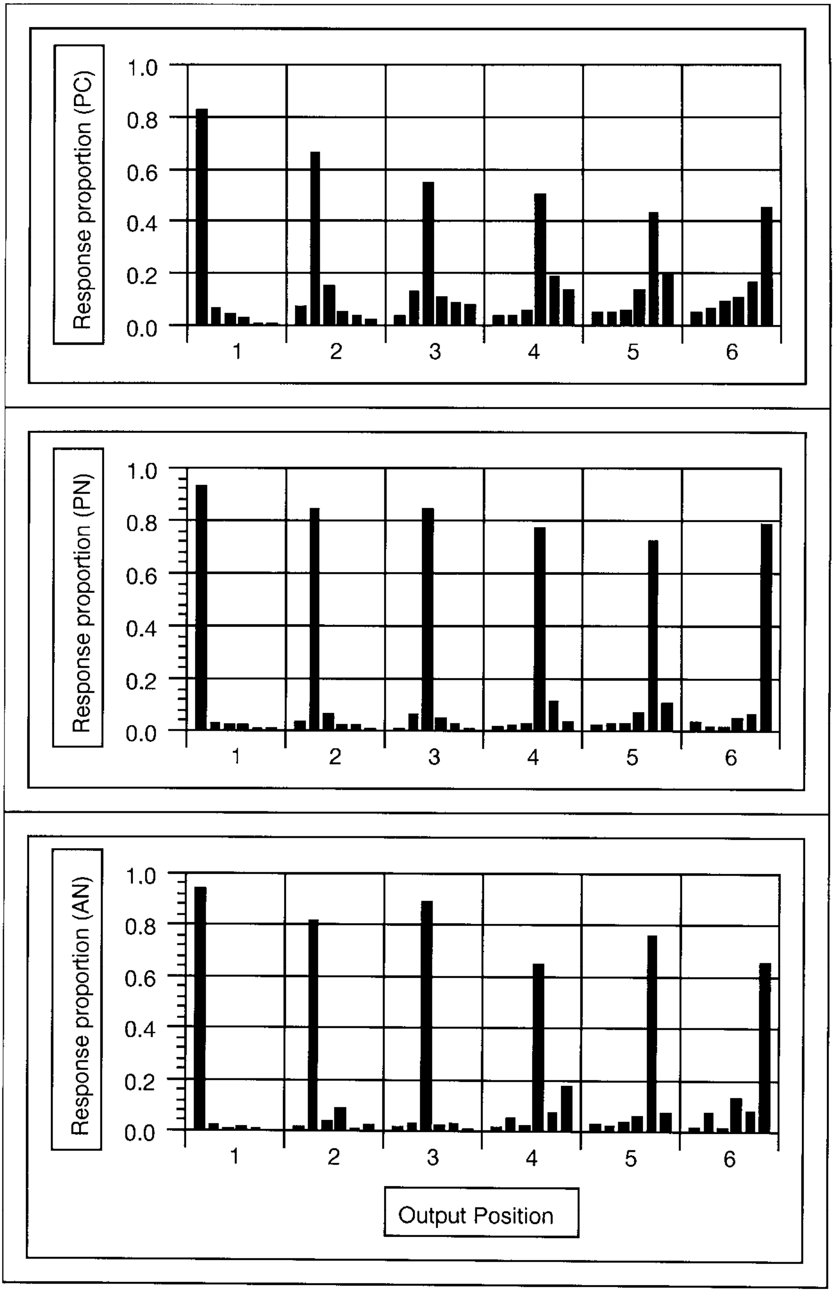

FIG. 2. Transposition gradients for each output position for confusable list type PC (upper panel), non-confusable list type PN (middle panel), and alternating list type AN (lower panel).

empirical constraints: the *locality constraint*, that transposed items tend to be localized around their input positions, and the *similarity constraint*, that a phonologically confusable item is more likely to substitute for another confusable item than a non-confusable item. These constraints appear to interact, in the sense that there is non-additive effect of phonological similarity in confusable versus non-confusable transposition gradients, particularly for the first output position.

Repeat Errors

Every erroneous response that occurred more than once in a report was classified as a repeat error. Repeat errors comprised typically around 16% of transpositions for all list types, their frequency thus being little affected by phonological similarity. Repeated items also tended to be far apart in a report (3.4 output positions on average), such that most repeat errors occurred on Output Positions 5 and 6, and were usually repeats of items that were reported correctly in Input Positions 1 or 2 (explaining the exceptions to the locality constraint found in later output positions for non-confusable lists).

The significance of this distribution of repeat errors can be qualified by considering a simple model of serial recall. Subjects who fail to recall an item correctly at a given output position might guess randomly from the six possibilities. Simulations of such a simple model, fitted to overall error rates, show about 84% of errors to be repeat errors for non-confusable lists and 64% for confusable lists. Not only are both simulation results far in excess of experimental results, but they also show a frequency of repeat errors that decreases with increasing error rate, which is in contradiction to the data. Simulations also give a mean distance between repeated items of 2.2 output positions in both cases, considerably smaller than in the data. Different percentages of repeat errors would be expected if subjects' guesses were biased towards neighbouring list items, as the locality constraint requires, but this biasing should lead to an even smaller mean distance between repeated items (i.e. less than 2.2).

In summary, repeat errors in the present experiment can be characterized by a third empirical constraint, the *repetition constraint*: Repeat errors are rare, and repeated responses are far apart. It suggests that people are reluctant to repeat a response made earlier in the same report (especially when they are aware that lists contain no repeats), and if they do, it is only likely after a number of intervening responses. Whether this reluctance reflects a conscious or unconscious process, it may be viewed as some general form of suppression of previous responses (e.g. Houghton, 1990). Obviously, the fact that significant numbers of repeat errors do occur suggests that this *response suppression* cannot be perfect. Rather, it seems to weaken over the course of report, explaining why repeated items are generally far apart. A similar explanation might be applied to the Von Ranschburg effect, where the second occurrence of a repeated item is often omitted.

Relative Errors

Every erroneous response that followed a response of the same item it followed in the list was classified as a relative error. The mean percentage of adjacent transposition errors that were relative errors was then calculated for non-confusable and confusable lists, for the 33 subjects that made at least one pair of adjacent transpositions in both list types. A related test of weighted differences between log-odds gave no evidence for a greater proportion of relative errors for non-confusable lists (*mean* = 0.23, *SD* = 0.22) than for confusable lists (*mean* = 0.22, *SD* = 0.15), $n = 33$, $Z = 0.08$, $p = 0.94$. Furthermore, with only six items per list, an independent guessing model will predict relative errors to comprise 20% of adjacent transpositions (given that the erroneous second response can only be one of five possible items). Any chaining model, however, would surely predict more than a chance level of relative errors, because the first erroneous response will increase the chance of the second response being its successor in the list. Yet weighted log-odds scores showed no reliable increase in the proportion of relative errors compared with a 20% chance level for either non-confusable lists, $Z = 0.23$, $p = 0.82$, or confusable lists, $Z = 0.25$, $p = 0.80$. It is worth noting that even if the proportion of relative errors for non-confusable lists had been closer to 30%, the null hypothesis could still not have been rejected at the 0.05 level (power = 99%).

In summary, the frequency of relative errors gives no evidence for an effect of similarity at cuing and, furthermore, fails to support the prediction of any chaining model that the frequency of relative errors should be above chance.

EXPERIMENT 2

Experiment 2 reexamined the alternating and non-confusable lists in Experiment 1, but with different sets of consonants, equated as far as possible for their predictability. Subjects were also split into two groups on the basis of a simple pretest of their span: High-span subjects were subsequently tested with lists of seven consonants; low-span subjects were tested with lists of six consonants. This division was to reduce the variability amongst subjects found in Experiment 1 and hence increase the power of the tests employed. The main tests were whether errors in recall of non-confusable items, particularly first-in-report errors, were more likely in recall of alternating lists than non-confusable lists, as is predicted by chaining models.

Method

Subjects. The 24 subjects were drawn mainly from the APU Subject Panel, and the rest were APU members; 10 were male, 14 were female, and their mean age was 28.

Materials. The consonants used in each list are shown in Table 3, together with their predictability. Consonants were less predictable on average than in Experiment 1, and importantly, their predictability was more closely balanced across different list types. The same set of 6 consonants was used to generate all 6-item *AC* and *AN* lists. Due to the odd number of items in 7-item lists, *AC* and *AN* differed in one consonant (an extra non-confusable consonant in *AN*; an extra confusable consonant in *AC*). Otherwise, the lists obeyed the same constraints as in Experiment 1.

Procedure. After subjects had read the instructions on the VDU, they were given a short "span" pretest consisting of eight lists, including examples of each list type and each list length. Subjects were divided into the *low-span group*, for those who made 14 or more errors on the pretest, and the *high-span group*, for those who made fewer than 14 errors (this criterion being determined from a

TABLE 3
List Types and Consonant Predictabilities in Experiment 2

List Length	*List Type*	*List Structure*[a]	*Letter Set (Example List)*	*Predictability*		*No. of Lists*
				mean	*SD*	
6	PN	NNNNNN	VRMJQH	0.89	0.17	18
	AC	CNCNCN	DQTJVM	0.70	0.23	18
	AN	NCNCNC	MTJVQD	0.70	0.23	18
7	PN	NNNNNNN	HJMRYQV	1.03	0.28	21
	AC	CNCNCNC	TJBMVQD	0.88	0.23	21
	AN	NCNCNCN	QDMVYTJ	1.14	0.24	21

[a] C = confusable item, N = non-confusable item.

pilot study). Every subject then attempted recall of three blocks of lists, containing either 18 lists of 6 consonants (the low-span group) or 21 lists of 7 consonants (the high-span group). Presentation rate was slightly slower than in Experiment 1, with 600 msec per item and 150 msec between items. Grouping instructions for 6-item lists were to use *3–3* grouping; for 7-item lists, instructions were to use *3–4* grouping. Otherwise, the instructions and remaining procedure were identical to Experiment 1.

Results and Discussion

In brief, error position curves for 6- and 7-item lists differed in that the troughs of the sawteeth for alternating lists lay almost perfectly on top of the non-confusable curve for the high-span group, whereas the low-span group exhibited sawteeth whose troughs sometimes lay slightly above this curve. However, conditional error position curves showed exactly the same pattern for both list lengths: The proportion of first-in-report errors in recall of non-confusable items in alternating lists never differed significantly from the proportion in non-confusable lists.

Overall Performance

The low-span group contained 13 subjects (pretest errors: *mean* = 20, *SD* = 5.6). They recalled correctly approximately 56% of *PN* lists, 37% of *AC* lists, and 42% of *AN* lists. Of the 1,006 errors in total, 21% were omissions and 5% were intrusions. Subjects in the low-span group made about the same mean number of errors per non-confusable list (*mean* = 1.1, *SD* = 0.85) as did subjects in Experiment 1. This probably reflected the slower presentation compensating for the preselection of low-span subjects and the less predictable nature of the lists. However, these factors may have increased the incidence of omissions. As intended, though, there was less variation in error rates than in Experiment 1.

The high-span group contained 11 subjects (pretest errors: *mean* = 7.4, *SD* = 4.7). They recalled correctly approximately 60% of *PN* lists, 46% of *AN* lists (with three confusable items), and 36% of *AC* lists (with four confusable items). Despite the extra item, subjects in the high-span group made roughly the same mean number of errors per non-confusable list (*mean* = 0.99, *SD* = 0.76) as did the low-span group. The 950 total errors included approximately 17% omissions and 6% intrusions. The greater proportion of omissions for this group, compared with Experiment 1, is probably because more subjects were facing a list length close to their span.

Error Position Curves

Error position curves are shown for each list length in the two panels in Figure 3. The general pattern resembles that from Experiment 1, although for lists of length 6, the sawteeth tend to lie above the non-confusable curve. Indeed, the difference between weighted log-odds scores on Output Positions 2, 5, and 6 in alternating and non-confusable 6-item lists approached significance, $N = 13$, $Z > 1.7$, $p < 0.05$ in all cases. For 7-item lists, however, even collapsing across non-confusable positions, there was no such significant difference, $N = 11$, $Z = 1.1$, $p = 0.28$.

Conditional Error Position Curves

The subset of first-in-report errors was also examined by output position, as shown in the conditional error position curves in Figure 4. These graphs show how often an error occurred at each output position, given that items in all previous positions had been recalled correctly (hence each report can only contribute at most one data point). The points are calculated from means of weighted log-odds scores.

On conditionalizing, the differences between error proportions for non-confusable items in alternating versus non-confusable curves were reduced. Even collapsing across non-confusable positions, weighted log-odds scores gave no evidence for a greater proportion of first-in-report errors for non-confusable items in alternating lists than in non-confusable lists, for either 6-item lists, $N = 13$, $Z = 1.0$, $p = 0.30$, or 7-item lists, $N = 11$, $Z = 0.02$, $p > 0.99$. At the same time, however, tests were sensitive enough to show that proportions on confusable positions in alternating lists were significantly greater than on corresponding non-confusable positions in non-confusable lists for all positions in 6-item lists, $Z > 2.0$, $p < 0.05$, except the last, $Z = 0.35$, $p = 0.72$, and all positions in 7-item lists, $Z > 1.8$, $p < 0.05$, except the first, $Z = 1.2$, $p = 0.22$. Taken together, these results are clearly incompatible with any prediction of chaining models for an effect of similarity at cuing, suggesting its effects are restricted to retrieval only.

Conditional error position curves were generally lower and flatter than unconditional error position curves, lacking the prolonged primacy effects found in the latter. The proportion of conditional errors on the last position for 7-item non-confusable lists (*mean* = 11%), for example, was about half the proportion of unconditional errors on that position (*mean* = 20%), a difference that proved significant, $N = 11$, $Z = 2.6$, $p < 0.01$. If the conditional probability of making an error on Output Positions 2–6 is less than the unconditional probability, the percentage of errors across output positions in error position curves cannot be independent. Note that dependency between errors across output positions restricts interpretation of the linear contrasts performed in Experiment 1 on unconditional error position curves, because large numbers of errors on one position

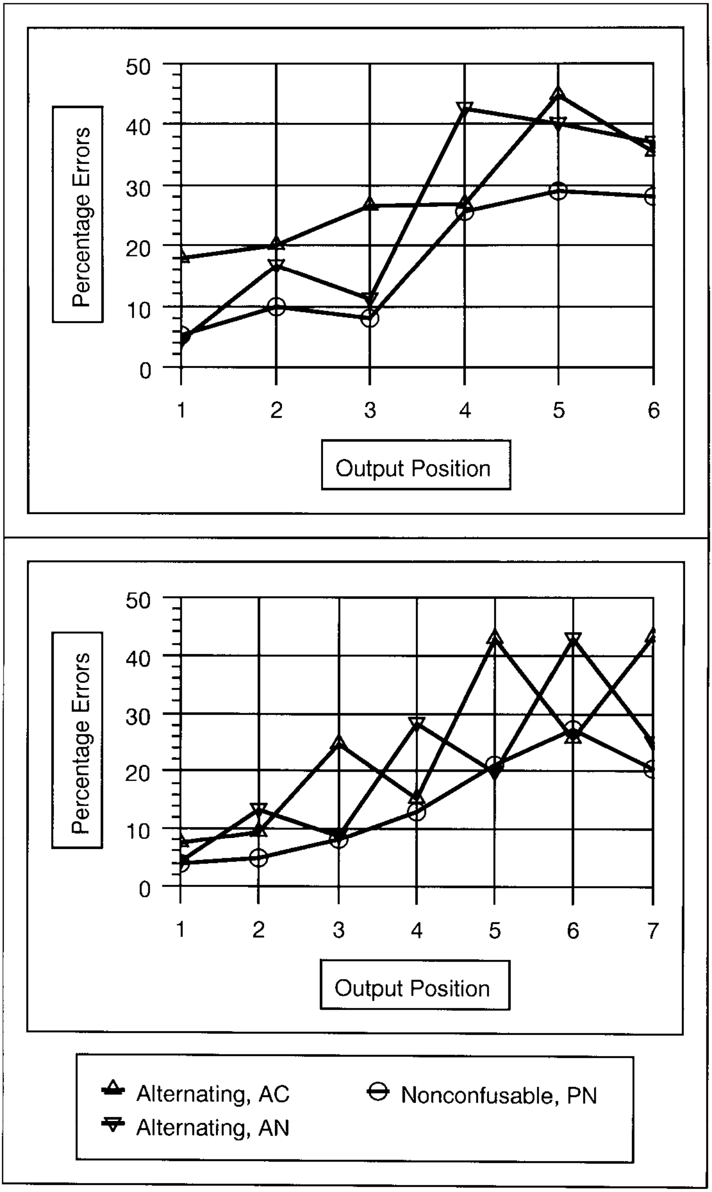

FIG. 3. Error position curves for each list type of length 6 (upper panel) and length 7 (lower panel) in Experiment 2.

will be associated with large numbers on other positions. (Indeed, such dependency between errors on different output positions has often been overlooked in numerous other short-term memory studies, which treat serial position as a factor with independent levels). However, the fact that the sawteeth remain when only first-in-report errors

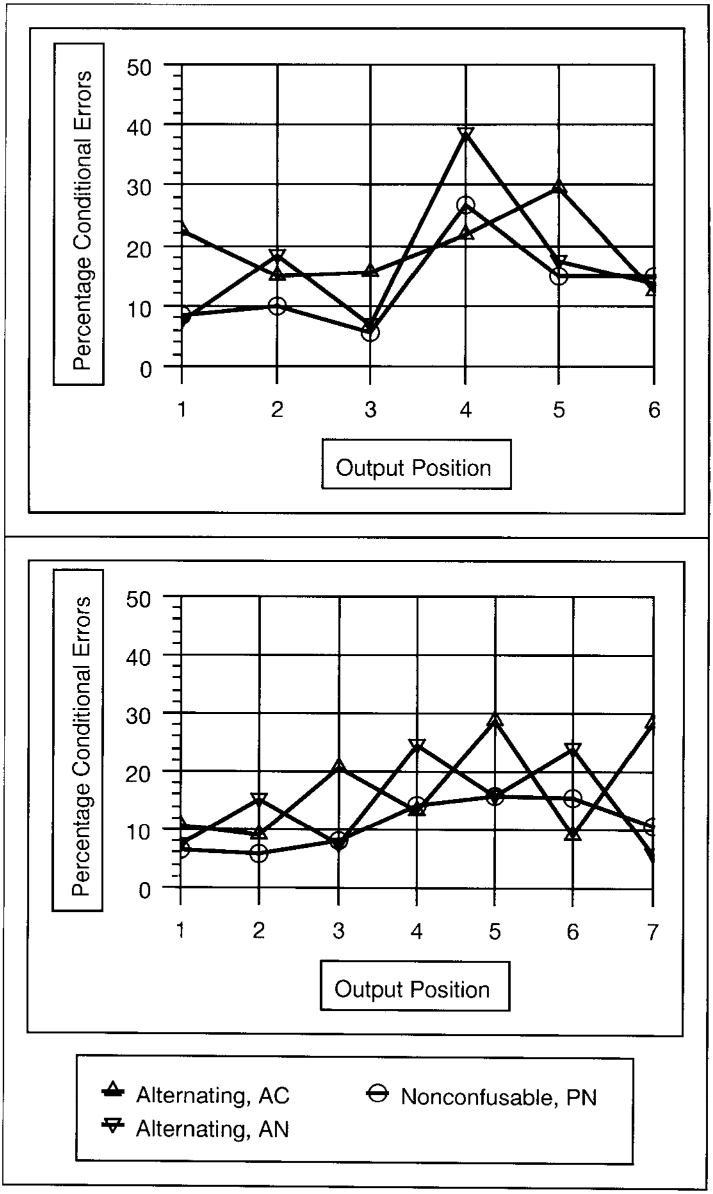

FIG. 4. Conditional error position curves for each list type of length 6 (upper panel) and length 7 (lower panel) in Experiment 2 (percentages are calculated from means of weighted log-odds scores).

are considered reinforces the conclusion that correct recall is generally harder for a confusable than a non-confusable item.

There are several ways in which an interdependence between responses might come about. Firstly, response suppression means that the set of possible responses becomes smaller with each item recalled (i.e. response selection is being made virtually without replacement). Hence, if early items are recalled correctly, then later items are more likely to be recalled correctly. A further cause of dependence between errors across output positions may simply be that subjects are disturbed when they experience difficulty in recall. Uncertainty or hesitation over a response may disrupt recall of following items, especially if recall is very sensitive to output delay (e.g. Cowan et al., 1992). Indeed, some subjects, particularly those in the low-span group, often ‘‘gave up’’ on remaining items as soon as recall faltered (supported in the data by a monotonic increase in omissions with output position). This ‘‘knock-on’’ effect will be greater for more difficult lists, which might explain why the sawteeth for 6-item alternating lists tended to lie above the non-confusable curve. For *AC* lists, for example, where the first item is confusable, any difficulty in recalling this item might have strong knock-on effects on recall of the rest of the list—at least, stronger than for *PN* or even *AN* lists. The potential for knock-on effects emphasizes the importance of conditional analysis as a tool for the study of serial position effects (cf. Johnson, 1972). Note, however, that previous studies have used *transitional error probabilities*, or *transitional shift probabilities*, which are only conditionalized on the immediately preceding response. Given that response selection in the present experiments may be operating from a small vocabulary without replacement, first-in-report errors should actually be more accurate indices of dependencies in reports.

EXPERIMENT 3

Several researchers have suggested that item-item associations and chaining might play a more prominent role in the auditory modality (e.g. Penney, 1989; Drewnowski, 1980). Baddeley’s (1968) Experiment VI, which was a variation on Experiment V with auditory presentation, produced sawteeth that lay above the non-confusable curve and whose peaks appeared coincident with the upper confusable curve. However, Experiment VI also differed from Experiment V in employing positional rather than serial recall, where written reports were required to indicate serial order, but actual order of recall was unconstrained. As Metcalfe and Sharpe (1985) found no difference in performance between vocalized and auditorily presented material (both showed a modality advantage over silent, visual presentation), Experiment 3 extends the serial recall conditions of Experiments 1 and 2 to the auditory modality via vocalized presentation and spoken recall.

Method

Subjects. Twenty-nine subjects from the APU Subject Panel—7 male and 22 female, with a mean age of 25—were tested.

Materials and Procedure. Stimuli were as in Experiment 2. Subjects were instructed to vocalize each letter as it appeared, and recall was spoken, unpaced, and recorded by the experimenter. Subjects were told to say ‘‘blank’’ if they could not remember a particular letter. The same presentation rate of one item every 0.75 sec was used as in Experiment 2. The relatively slow rate makes any separate ‘‘streaming’’ of confusable and non-confusable items (e.g. Jones, 1992) unlikely. Grouping was also made explicit by inserting a 0.75-sec pause between the third and fourth letter. This was to reinforce consistent grouping strategies further across subjects. Otherwise, the procedure was identical to Experiment 2.

Results and Discussion

In brief, the same pattern of results arose as in Experiment 2. Whereas the proportions of errors in recall of non-confusable items in alternating lists were sometimes greater than in non-confusable lists, for both list lengths in this case, the proportions of first-in-report errors for such items did not differ significantly.

Overall Performance

The low-span group contained 15 subjects (pretest errors: *mean* = 24, *SD* = 7.2). They recalled correctly approximately 68% of *PN* lists, 41% of *AC* lists, and 46% of *AN* lists. They made fewer errors on average per non-confusable list (*mean* = 0.79, *SD* = 0.66) than did subjects in either Experiment 1 or Experiment 2. This probably reflects the modality advantage of vocalizing items during presentation and the effect of explicit grouping. Of the 1000 errors in total, 10% were omissions and 7% were intrusions.

The high-span group contained 14 subjects (pretest errors: *mean* = 6.9, *SD* = 4.6). They recalled correctly approximately 58% of *PN* lists, 38% of *AN* lists (with three confusable items), and 19% of *AC* lists (with four confusable items). Probably due to the extra item, subjects made slightly more errors per non-confusable list (*mean* = 0.97, *SD* = 0.66) than did those in the low-span group. The 1433 errors in total included approximately 7% omissions and 5% intrusions.

Error Position Curves

Error position curves for the two list lengths are shown in Figure 5. Strong grouping effects were seen in the non-confusable curve, with large recency effects at the end of groups. Sawteeth for alternating lists were present at both list lengths, but for *AC* lists they tended to lie above the non-confusable curves, much like the pattern found in Baddeley's (1968) Experiment VI and for the 6-item lists in Experiment 2. Though the differences did not approach significance in the case of *AN*, weighted log-odds scores showed that more errors were made in recall of all non-confusable items in *AC* lists than in non-confusable lists for 6-item lists, $N = 15$, $Z > 1.6$, $p < 0.10$ in all cases, and on Output Position 6 for 7-item lists, $N = 14$, $Z = 3.2$, $p < 0.005$.

Conditional Error Position Curves

Conditional error position curves for the two list lengths are shown in Figure 6. In addition to flattening the curves, conditionalizing also removed any reliable differences in error proportions on non-confusable positions in alternating versus non-confusable lists. Even collapsing across non-confusable positions, tests of weighted, log-odds scores did not show a significantly greater proportion of first-in-report errors for non-confusable

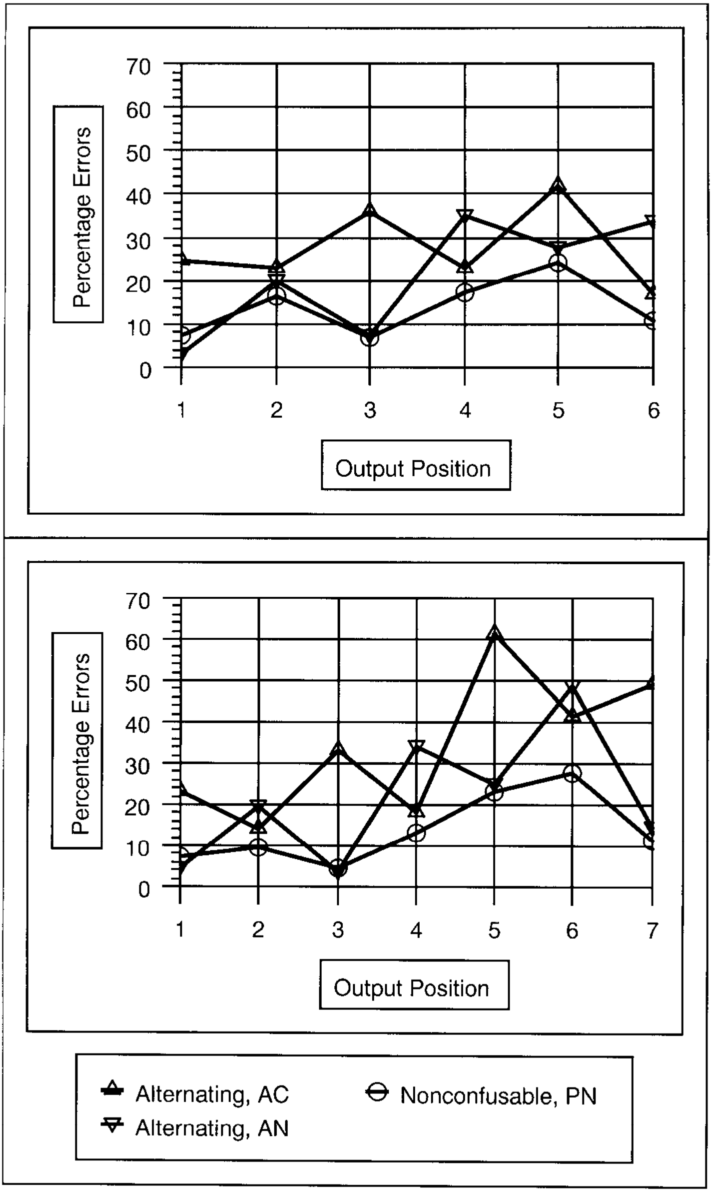

FIG. 5. Error position curves for each list type of length 6 (upper panel) and length 7 (lower panel) in Experiment 3.

items in alternating than non-confusable lists, for either 6-item lists, $N = 15$, $Z = 1.2$, $p = 0.24$, or 7-item lists, $N = 14$, $Z = 0.1$, $p = 0.92$. On the other hand, there was a significantly greater proportion of first-in-report errors on confusable positions in alternating lists, compared with corresponding non-confusable positions in non-confusable

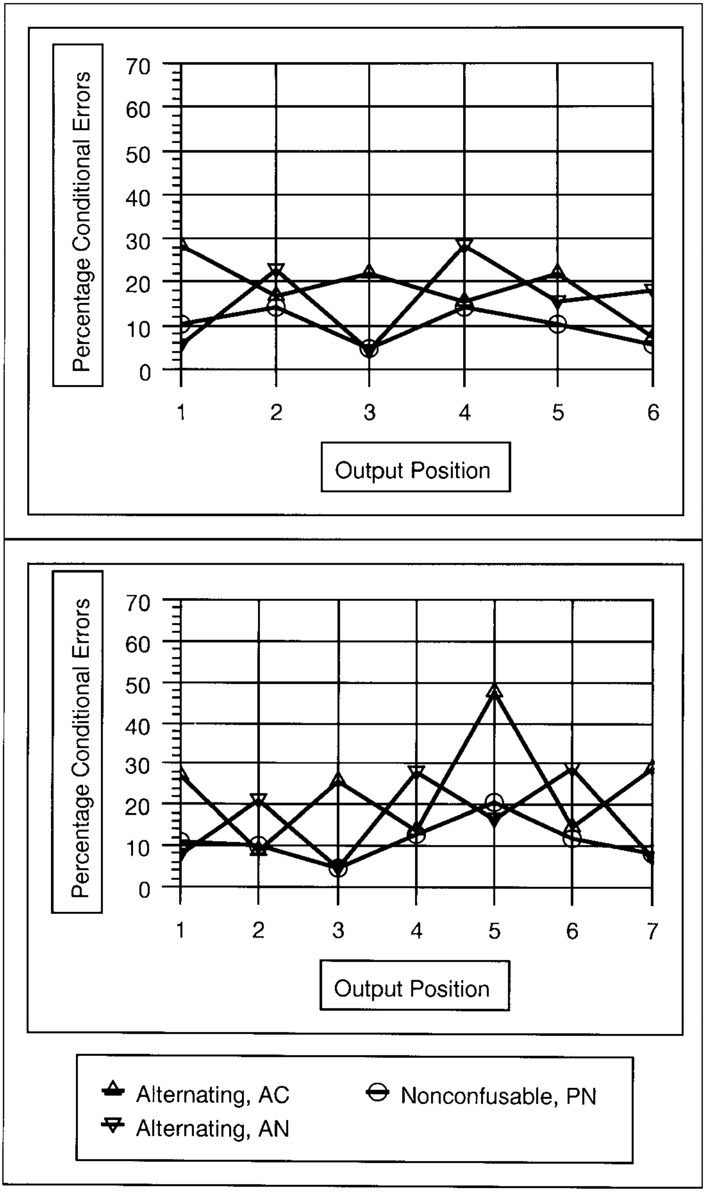

FIG. 6. Conditional error position curves for each list type of length 6 (upper panel) and length 7 (lower panel) in Experiment 3 (percentages are calculated from means of weighted log-odds scores).

lists, for both 6-item lists, $Z > 2.6$, $p < 0.01$, in all cases, and 7-item lists, $Z > 3.2$, $p < 0.005$ in all cases.

These results support the conclusions of Experiment 2. Although confusable items sometimes increase the chance of erring in recall of subsequent non-confusable items (as evident from error position curves), this appears to be the case only when those confusable items have themselves failed to be recalled correctly. In other words, any effect confusable items have on non-confusable items is not in virtue of phonological similarity at cuing (as evident from conditional error position curves). Rather, their effect on non-confusable items may be due to the general knock-on effect in serial recall when recall of an item becomes difficult and hence error-prone, as described in Experiment 2. Knock-on effects might be equally relevant to Baddeley's positional recall, especially if subjects default to forward recall when a recall order is not specified. Moreover, knock-on effects are likely to be stronger in spoken than written report, because previous responses cannot be reperceived and because of the greater difficulty in reordering responses. Consequently, although error position curves in Experiment 3 and Baddeley's (1968) Experiment VI differ slightly from those with visual presentation, this is not necessarily evidence for chaining in the auditory modality.

GENERAL DISCUSSION

The pattern of results emerging from the present experiments consistently fails to provide any support for chaining theories. In fact, they are difficult to reconcile with any current chaining model.

Error position curves for alternating lists show prominent sawteeth, the peaks of which reflect significantly greater numbers of errors in recall of confusable than non-confusable items. Although chaining models can be constructed that are compatible with this sawtooth pattern, they remain unable to explain the fact that, whether unconditional or conditional error probabilities are examined, the presence of confusable items in a list generally has no detectable influence on the probability of correctly recalling surrounding non-confusable items.

Chaining models such as Richman and Simon's (1994) EPAM model can account for the sawtooth pattern via phonemic confusions in retrieval. However, once a response is chosen, the cue for the next item becomes the corresponding "letter chunk". The greater number of errors on preceding confusable positions means that a chunked cue used for retrieval of the next non-confusable item is more often incorrect in recall of alternating lists than in recall of non-confusable lists. This should result in more errors in recall of non-confusable items in alternating lists than in non-confusable lists. A similar prediction would be made by models with context-sensitive token representations (e.g. Wickelgren, 1969), in which cuing with incorrect tokens following frequent errors in recall of a confusable item would likewise mean that the troughs of the sawteeth should always lie above the non-confusable curve. However, present results are in agreement with those of Baddeley's (1968) Experiment V, showing that the troughs in unconditional error position curves are more often coincident with the non-confusable curve. In fact, over Experiments 2 and 3, the troughs only diverge significantly from the lower curves on about 27% of occasions, mostly for later positions. Moreover, there is an alternative, non-chaining explanation for this in terms of "knock-on" effects, given in the discussion of Experiment 2. Thus, inasmuch as a simple chaining model predicts a general detrimental effect of confusable items on recall of subsequent non-confusable items, any effect must be small and may well have other explanations.

One way to square chaining models with the bounded sawteeth in error position curves is to consider models that allow an effect of similarity at cuing as well as retrieval. This is possible with models that chain along phonological representations, including that of Wickelgren (1965), as well as more sophisticated recurrent neural networks, such as Kleinfeld (1987) and Jordan (1986). They predict that a response following a confusable cue, whether that cue is the immediately preceding response or some compound of previous responses, is less likely to be in the correct order relative to the previous response than a response following a non-confusable cue. As shown in the Appendix, this allows the possibility that no more errors are made in recalling non-confusable items in alternating lists than in non-confusable lists, by virtue of the fact that erroneous, confusable cues reduce the risk of further errors and hence increase the chance of recovering from an error.

Evidence contrary to such models was found in the conditional error position curves in Experiments 2 and 3. When all previous responses are correct, an effect of similarity at cuing should result in more errors in recalling a non-confusable item when it followed a confusable item in alternating lists than when it followed another non-confusable item in non-confusable lists. However, the fact that troughs in the sawtooth conditional error position curves were always coincident with non-confusable curves showed that this clearly is not the case: The proportions of first-in-report errors in recall of non-confusable items in alternating lists never differed significantly from the proportion in non-confusable lists. The proportions on confusable positions, meanwhile, were nearly always significantly greater than on corresponding non-confusable positions in non-confusable lists. Thus there is again a consistent effect of similarity at retrieval, but no evidence for an effect at cuing. These results are troublesome for Murdock's TODAM model (Lewandowsky & Murdock, 1989). If phonological similarity is represented in TODAM's item vectors, the most natural way to represent phonological similarity in distributed memory models, then there should be an effect of similarity at cuing. (TODAM's selective cuing mechanism is not relevant to conditional error position curves, because the cue is always correct in these curves.) TODAM's general inability to simulate the Baddeley (1968) data was confirmed by Baddeley, Papagno, and Norris (1991). The possibility remains that TODAM could keep its item representations as random vectors, and model phonological similarity as affecting only retrieval, or the "deblurring" of the results of chaining. However, such a non-phonological chaining model, even with selective cuing, then faces the problem of the near-coincidence of troughs of alternating curves with non-confusable curves in conventional error position curves, as discussed earlier. These criticisms would also appear to apply to Murdock's TODAM 2 model (Murdock, 1993), which still employs an element of chaining.

Finally, converging evidence against chaining models arises from examining the frequency of relative errors. Chaining models have to predict that, once an error has been made, the erroneous response will provide an imperfect cue for the next response. Other things being equal, this will increase the chance of the next response also being an error. Specifically, it should increase the chance that the subsequent response is the item that followed the erroneous response in the list. This means that although the two responses will be in the correct relative order, both will be in the wrong position. Further, models that chain along phonological representations predict a greater frequency of such relative errors in recall of non-confusable lists than confusable lists, because non-confusable cues are more likely than confusable cues to lead to a relative error. However, Experiment 1 showed no evidence whatsoever of any such difference in the proportion of adjacent transpositions that were relative errors. Moreover, although exact proportions are nearly impossible to calculate analytically (especially with additional factors such as response suppression), the fact that the proportion of adjacent transpositions that were relative errors in recall of either confusable or non-confusable lists was not significantly greater than 20%—the frequency predicted by random guessing—is surely at variance with any chaining model.

In summary, all chaining models described above predict an effect of confusable items on the recall of subsequent non-confusable items, whether that is by virtue of phonological similarity between confusable cues or by virtue of the increased number of errors in recall of preceding confusable items. However, the data tell a different story: Whether or not previous responses were correct, confusability of items had no consistent, detectable effect on the probability of recalling a subsequent non-confusable item. A similar conclusion was also reached by Bjork and Healy in 1974: ". . . it appears that the presence of two acoustically similar items in the same to-be-remembered stimulus does not increase the loss of order information for all letters in the stimulus string but rather produces rapid loss of order information specific to the two similar letters" (p. 91). Knock-on effects aside, this statement appears to apply whether stimuli were presented visually in silence (Experiment 2) or vocalized (Experiment 3), which is also problematic for more general theories that propose an element of chaining in the auditory modality (e.g. Drewnowski, 1980; Penney, 1989).

If item–item associations and chaining are inappropriate means of storing and retrieving order information, what other options are available? One possibility will be highlighted by describing the *primacy model* (Norris, Page, & Baddeley, 1994; Page & Norris, submitted). As well as giving very close fits to error position curves from Baddeley's original experiment, an achievement not met by any other model (Baddeley et al., 1991; Burgess & Hitch, 1992), it also accounts for further, detailed aspects of the present data, which appear problematic for alternative, non-chaining accounts. (The scope of the primacy model is not, of course, limited to accounting for the present data; however, a complete account would extend well beyond the present remit.)

The Primacy Model

In the primacy model, serial recall is supported not by associations between items but by a *primary gradient* of activation across item representations. This approach, in which a primacy gradient stores not item–item or position–item information, but order itself, has been suggested previously (e.g. Grossberg, 1978), although never before developed into a detailed model of serial recall. One way to imagine how such an activation gradient might arise is by means of association between each item and a representation of the start of the list. The strength of association between the start of the list and each successive item decreases across input position. When pre-list context is used to cue recall, a node representing each item in the list is activated in proportion to the strength of the item's original association. This results in a primacy gradient of activation across nodes, as shown in Figure 7. All nodes are activated simultaneously, and the activation of each node is a fixed ratio of the activation of the node representing the preceding item.

Recall of the list begins by choosing the node with the highest activation and then suppressing activation in that node to prevent that item's being chosen again. For non-confusable items, the chosen node corresponds to the response made (for confusable items, the final response is affected by a second stage described later). With no noise in the system, recall would be perfect. Errors are introduced into the system by adding zero-mean Gaussian noise independently to each node. The variance of this noise and the ratio of successive activation levels are the two basic parameters of the model.

A simulation of a two-parameter version of the primacy model for the non-confusable lists in Experiment 1, fitted by a modified NDMA algorithm (Caprile & Girosi, 1990), produces the error position curve in Figure 8, shown together with the data. The model gives the asymmetrically bowed curve, with a recency effect restricted to the last item, that is characteristic of immediate serial recall of visually presented items. Although this two-parameter model does not produce any omissions, intrusions, or repeated responses, the fit is surprisingly good: The residual, root mean square (RMS) error is only 3.3%. The main discrepancies occur on Output Positions 3, 4, and 6. The former two are most probably due to grouping effects in the data; the latter occurs because the model predicts stronger last-item recency than is found in the data. However, by allowing the repetition of responses, the model can produce a recency effect that is more in line with the data. If response suppression is allowed to wear off slowly during recall, through the addition of a third parameter, repeated responses become possible, though generally after a number of

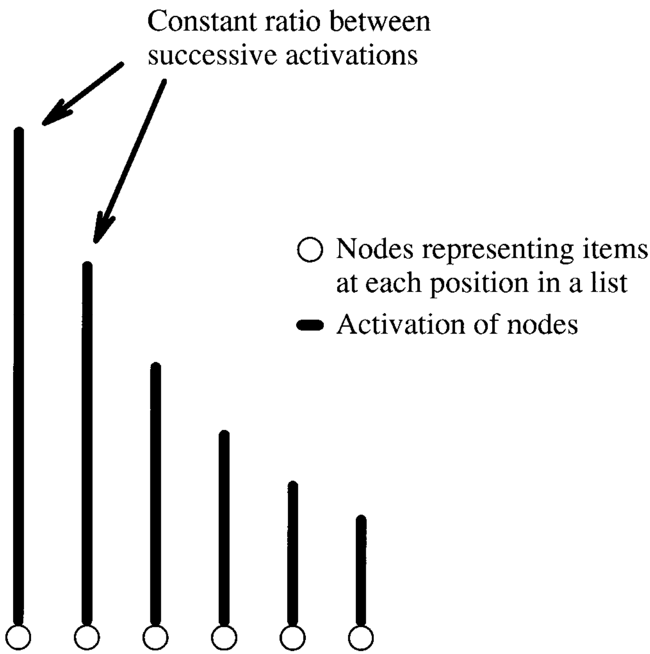

FIG. 7. Schematic primacy gradient in the primacy model.

intervening responses, as required by the repetition constraint. In fact, in agreement with the data, repeat errors occur towards the end of the report, being repeats of items reported correctly at early positions. This diminishes the recency effect, and the RMS error can be reduced to 1.4%.

It is important to realize that the two basic parameters serve only to help the model give a good quantitative fit to the data. The central qualitative predictions of the model are a straightforward consequence of the combined effects of the primacy gradient in association with response suppression. But just what is it about this model that enables it to give such a good explanation of the basic data from serial recall?

In recalling an item, that item has to compete with all of the remaining items that have not yet been recalled. Items early on in the list have little trouble winning this competition. Although these items have the greatest number of potential competitors, the fact that the gradient is determined by a ratio rule means that there is a larger difference in activation between the node for an early item and the node for the immediately following item than there is between nodes for a later item and its successor. Larger absolute differences in activation are harder to bridge by additive noise, so early items are therefore more easily recalled in their correct position. For recall of items in the middle of the list, the difference in activation levels between respective nodes becomes less, while there is still a significant number of competitors. Towards the end of recall, however, the reduction in the number of competitors compensates for the poorer discriminability of later items. The error position curve, therefore, emerges as a consequence of a trade-off across output positions between discriminability and number of competitors.

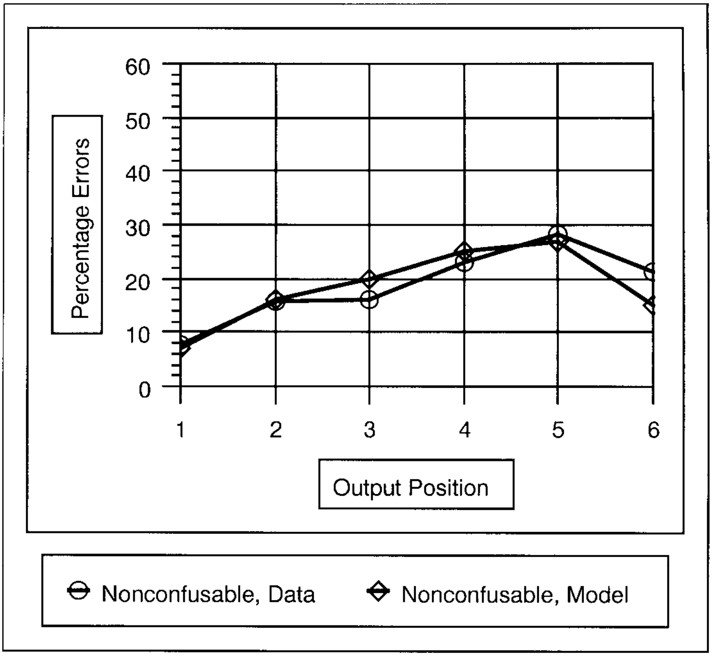

FIG. 8. Error position curves for non-confusable lists in Experiment 1: data, and simulation results from the primacy model.

One of the crucial features of the primacy model that enables it to give such an accurate account of transposition errors is that it exhibits a property referred to as *fill-in*. Fill-in arises because the primacy gradient ensures that when an item fails to be recalled in its correct position, the node representing that item remains highly activated and is very likely to be chosen for recall in the next position (explaining why the majority of errors are paired transpositions—i.e. the swapping of adjacent items, and not relative errors). It is the fill-in property that allows the model to satisfy the powerful locality constraint.

The effects of phonological similarity emerge at a second stage of processing. A second set of nodes, the output nodes, have activations determined by the similarity of the items they represent to the item that was chosen from the first stage. A confusable item chosen from the first stage will activate its representation in the second stage as well as activating, to a lesser extent, output nodes for similar items. Activation of all output nodes is also multiplied by the activation of corresponding input nodes in the first stage, superimposing another primacy influence over the activated output nodes. This produces the non-additivity of phonological similarity on transposition gradients, or the interaction between the locality and similarity constraints described in Experiment 1. Noisy selection of nodes then operates as in the first stage, with suppression of the node eventually chosen to represent the response.

The second stage requires the addition of two further parameters. One parameter represents the variance of Gaussian noise added to the second-stage activations; the second represents the extent to which a confusable item partially activates other confusable items (i.e. the degree of similarity between confusable items). Non-confusable items do not activate other items, and hence pass through the second stage unaffected. The second stage ensures that confusable items are most likely to transpose with other confusable items, meeting the similarity constraint. A greater number of second-stage competitors in recall of confusable lists also explains why curves for these lists lie above the peaks of alternating curves in Experiment 1 (even when predictability alone might suggest otherwise).

The four-parameter version of the model was fitted to the 6- and 7-item data from Experiment 2, as shown in Figure 9. The model clearly gives an excellent fit to the 7-item data. The RMS error over all 21 points is only 4.8%. Most importantly, the model reproduces the most theoretically significant aspect of these results: Recall of non-confusable items is not affected by the presence of confusable items in alternating lists. Note again that the fit to the data is exceptionally good, despite the fact that the four-parameter model does not account for omissions, intrusions, or repetitions. An extended version of the model that does make some allowance for such errors improves the fit to an RMS error of 2.8%. Clearly, the fit for the 6-item lists is much less impressive than that for the 7-item lists. This is most probably due to the knock-on effects in the data described earlier. However, the model still captures the underlying pattern of transposition errors, even for the 6-item lists. Whereas the fit of the model to the error position curves has an RMS error of 7.2%, the RMS error for fits to error position curves when only transpositions are considered is 2.2%.

The primacy model has several advantages over the Burgess and Hitch (1992) model. The Burgess and Hitch model employs a "context window" that moves across output

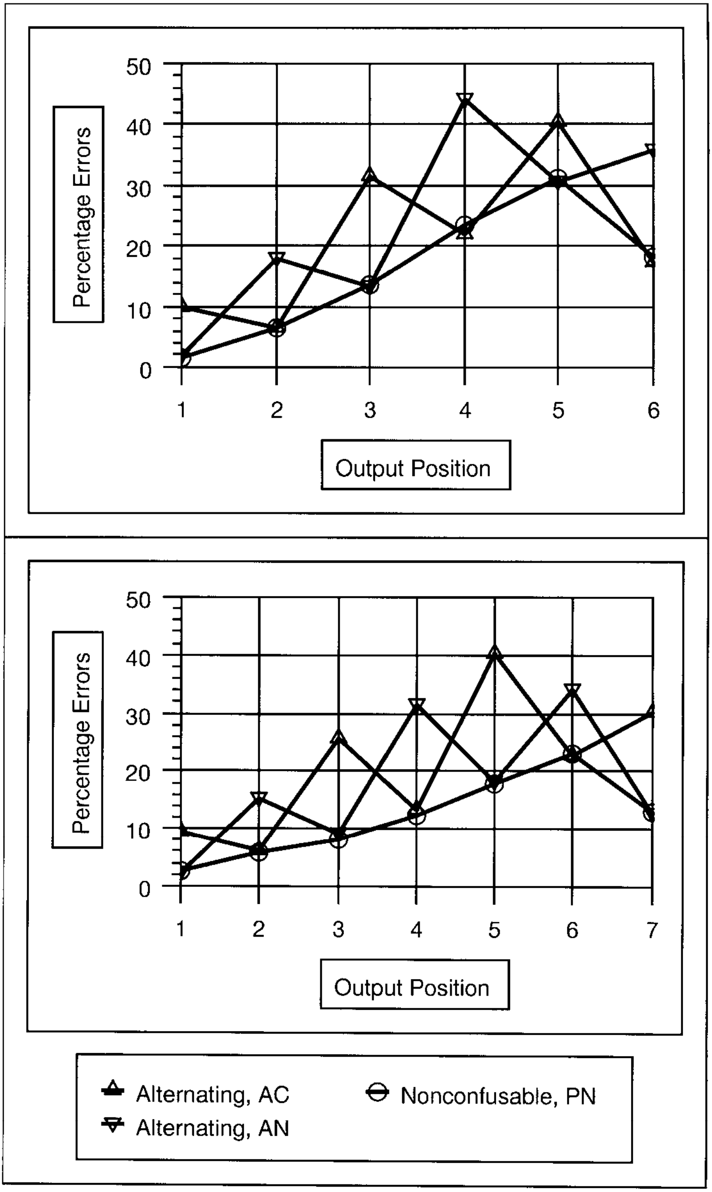

FIG. 9. Primacy model simulations of error position curves for each list type of length 6 (upper panel) and length 7 (lower panel) in Experiment 2.

positions during recall, cuing responses via associations between items and their positional context. One advantage of the primacy gradient is that it avoids the problem of how to reconstruct the context at recall. Using a moving context, each successive context vector must be reproduced; with a primacy gradient, only a single pre-list representation of the context need be reproduced, because this single pattern is sufficient to support recall of the entire list.

A second problem of the Burgess and Hitch model is that its fill-in properties are weak. Once item $i + 1$ has been recalled too early and the context window has moved to position $i + 1$, the symmetrical nature of the window means that, other things being equal, item i and item $i + 2$ are equally likely to follow. In fact, with the presence of greater decay of weights between context and early items, item i actually becomes less likely to follow than item $i + 2$. Futhermore, if item $i + 2$ does follow, the chance of filling in item i tends to get progressively smaller as the window moves on (as opposed to the primacy gradient model, where the chance of fill-in always gets progressively bigger). The lack of significant fill-in in the Burgess and Hitch model leads to weak recency effects and error patterns that are inconsistent with the data presented in Experiment 1.

A final problem of the Burgess and Hitch model arises when it tries to explain the lack of any effect of confusable items on recall of neighbouring non-confusable items. The problem is that the overlap in contextual cues will interact with the overlap of phonological representations of items (in the model's input phoneme layer), to produce greater contextual cuing of items when they are confusable than when they are non-confusable. This is because the weights common to the context and shared phonemes of confusable items get reinforced on more than one occasion. Thus, even when the context window is supposedly cuing recall of a non-confusable item, the confusable competitors in recall of alternating lists will provide stronger competition than non-confusable competitors in non-confusable lists, and consequently more errors will be expected on non-confusable positions in alternating than non-confusable curves. This would apply even to recall of the first non-confusable item in alternating lists, and yet none of Experiments 1 to 3 show any more errors on this position than the first position in non-confusable lists.

The main advantage of the primacy model over the perturbation theory of Lee and Estes (1978, 1981) is that the primacy model has an explicit recall process. Perturbation theory assumes random errors in the relative timings (phases) of the cyclic reactivation of item representations result in the perturbation of positions of items. The gradual accumulation of these perturbations produces the descending transposition gradients around an item's input position. Lee and Estes provide a mathematical description of the results of this process, allowing calculation of the expected distribution of errors over a large number of trials, but they do not have a model that can actually simulate individual reports (see Nairne & Neath, 1994, and Mewhort, Popham, & James, 1994, for a similar criticism of TODAM). This is because a fundamental assumption of the mathematical model is that items perturb independently. An item perturbing forward, for example, has no effect on the recall of other items, leading to impossible situations where more than one item is supposedly stored at the same position. However, the present experiments have shown just how important dependencies between responses are, given the sequential nature of the actual recall process. Though dependencies between responses will be weaker with a larger vocabulary or when repetition in lists is possible, the independency assumption of perturbation theory is clearly inadequate. The advantage of actually constructing computational models with non-linear, non-independent processes is that they can produce complex behaviour from relatively simple mechanisms (e.g. the surprisingly good serial position curves arising from a simple primacy gradient and suppression).

Conclusion

The present study has shown how detailed analysis of patterns of errors can shed considerable light on the nature of the mechanisms required in a successful model of immediate serial recall. The locality constraint shows that errors arise through mechanisms beyond random guessing. The repetition constraint suggests models must incorporate some form of response suppression during recall, which produces the interdependency between responses that is observed. The similarity constraint argues that phonological confusions arise through some form of response competition. The primacy model demonstrates appropriate sensitivity to all three of these constraints.

Chaining theories, in contrast, do not appear to present a viable mechanism for serial recall. Contrary to their predictions, confusable items seem to have little effect on recall of non-confusable items in the same list. Rather, competition held over a simple primacy gradient of activations of list items appears adequate for explaining the basic serial position effects. The specific phonological similarity effects then arise from confusions occurring at a second stage, subsequent to selection of a potential response for a given position. The primacy model appears to be the only model currently capable of providing an accurate, quantitative simulation of the present data. Furthermore, the model has been extended to include rehearsal, word-length and list-length effects (Page & Norris, submitted) and is being applied to effects of irrelevant speech and grouping in short-term memory.

REFERENCES

Baddeley, A.D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology*, *20*, 249–263.

Baddeley, A.D. (1971). Language habits, acoustic confusability, and immediate memory for redundant letter sequences. *Psychonomic Science*, *22*, 120–121.

Baddeley, A.D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A.D., Conrad, R. & Hull, J. (1965). Predictability and immediate memory for consonant sequences. *Quarterly Journal of Experimental Psychology*, *17*, 175–177.

Baddeley, A.D., & Ecob, J.R. (1970). Simultaneous acoustic and semantic coding in short-term memory. *Nature*, *227*, 288–289.

Baddeley, A.D., Papagno, C., & Norris, D.G. (1991). A sandwich for TODAM. In W.E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: In honor of Bennet B. Murdock* (pp. 175–194). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Burgess, N., & Hitch, G. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*. *31*, 429–460.

Caprile, B., & Girosi, F. (1990). *A nondeterministic minimization algorithm* (AI Memo number 104). Massachusetts Institute of Technology, Massachusetts, MA.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning*, *4*, 161–169.

Conrad, R., & Hull, A.J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 429–432.

Cowan, N., Day, L., Saults, J.S., Keller, T.A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*, 1–17.

Cox, D.R., & Snell, E.J. (1989). *Analysis of binary data* (2nd ed.). London: Chapman & Hall.

Drewnowski, A. (1980). Attributes and priorities in short-term recall: A new model of memory span. *Journal of Experimental Psychology: General*, *109*, 208–250.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*, New York: Dover.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Estes, W.K. (1972). An associative basis for coding and organisation in memory. In A.W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: V.H. Winston & Sons.

Frankish, C.R. (1974). *Organisational factors in short-term memory*. Unpublished doctoral dissertation, University of Cambridge.

Frick, R.W. (1995). Accepting the null hypothesis. *Memory & Cognition*, *23*, 132–138.

Grossberg, S. (1978). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, *17*, 199–219.

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

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.

Jahnke, J.C. (1969). The Ranschburg effect. *Psychological Review*, *76*, 592–605.

Jensen, A.R., & Rohwer, W.D. (1965). What is learned in serial learning? *Journal of Verbal Learning and Verbal Behaviour*, *4*, 62–72.

Johnson, N. (1972). Organisation and the concept of a memory code. In A.W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 125–160). Washington DC: V.H. Winston & Sons.

Jones, D. (1992). Objects, streams, and threads of auditory attention. In A.D. Baddeley and L. Weiskrantz (Eds.), *Attention: Selection, awareness and control (A tribute to Donald Broadbent)* (pp. 87–104). Oxford: Clarendon Press.

Jordan, M.I. (1986). *Serial order: A parallel distributed approach* (ICS Report 8604). San Diego: University of California, Institute for Cognitive Science.

Kleinfeld, D. (1987). Sequential state generation by model neural networks. *Proceedings of the National Academy of Sciences USA*, *83*, 9469–9473.

Lashley, K.S. (1951). The problem of serial order in behaviour. In L.A. Jeffress (Ed.), *Cerebral mechanisms in behaviour: The Hixon Symposium* (pp. 112–136). New York: John Wiley.

Lee, C.L. (1976). Short-term recall of repeated items and detection of repetitions in letter sequences. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 120–127.

Lee, C.L., & Estes, W.K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Memory*, *16*, 395–418.

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.

Metcalfe, J., & Sharpe, D. (1985). Ordering and reordering in the auditory and visual modalities. *Memory & Cognition*, *13*, 435–441.

Mewhort, D.J.K., Popham, D., & James, G. (1994). On serial recall: A critique of chaining in the Theory of Distributed Associative Memory. *Psychological Review*, *101*, 534–538.

Murdock, B.B. Jr. (1987). Serial order effects in a distributed memory model. In D.S. Gorfein & R.R. Hoffman (Eds.), *Learning and memory. The Ebbinghaus Centennial Conference* (pp. 819–827). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Murdock, B.B. Jr. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, *100*, 183–203.

Murdock, B.B. Jr., & vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology*, *74*, 137–143.

Nairne, J.S., & Neath, I. (1994). Critique of the retrieval/deblurring assumptions of the Theory of Distributed Associative Memory. *Psychological Review*, *101*, 528–533.

Norris, D.G., Page, M.P.A., & Baddeley, A.D. (1994, July). *Serial order: It's all in the representations*. Paper given at the International Conference on Working Memory, Cambridge.

Page, M.P.A., & Norris, D.G. (submitted). *The Primacy Model: A new model of immediate serial recall.*
Penney, C.G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition*, *17*, 398–422.
Posner, M.I., & Konick, A.F. (1966). On the role of interference in short-term retention. *Journal of Experimental Psychology*, *72*, 221–231.
Richman, H.B., & Simon, H.A. (1994). *EPAM simulations of short-term memory* (Complex Information Processing Working Paper #514). Carnegie-Mellon University, Department of Psychology.
Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, *21*, 137–147.
Shiffrin, R., & Cook, J. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behaviour*, *17*, 189–218.
Slamecka, N. (1967). Serial learning and order information. *Journal of Experimental Psychology*, *74*, 62–66.
Solso, R.L., & Juel, C.L. (1980). Positional frequency and versatility of bigrams for two-through nine-letter english words. *Behaviour Research Methods and Instrumentation*, *12*, 297–343.
Taylor, J.G. (1991). Neural network capacity for temporal sequence storage. *International Journal of Neural Systems*, *2*, 47–54.
Watkins, M.J., Watkins, O.C., & Crowder, R. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behaviour*, *13*, 430–447.
Wickelgren, W.A. (1965). Short-term memory for phonemically similar lists. *American Journal of Psychology*, *78*, 567–574.
Wickelgren, W.A. (1966). Associative intrusions in short-term recall. *Journal of Experimental Psychology*, *72*, 853–858.
Wickelgren, W.A. (1969). Context-sensitive coding, associative memory and serial order in (speech) behaviour. *Psychological Review*, *76*, 1–15.
Young, R. (1968). Serial learning. In T.R. Dixon & D.L. Horton (Eds.), *Verbal behaviour and general behaviour theory* (pp. 122–148). Englewood Cliffs, NJ: Prentice-Hall.

Manscript received 14 November 1994
Accepted revision received 30 April 1995

APPENDIX

Chaining Models and a Conditional Analysis of Errors

The results of sawtooth curves for alternating lists being contained within curves for pure lists in Baddeley's 1968 Experiment V are not actually sufficient to rule out effects of phonological similarity of one item on its successors (as entailed by models that chain along phonological representations). This is because a Bayesian probabilistic analysis of errors reveals the situation to be potentially more complex than is apparent simply from error position curves.

Sawteeth in Alternating Lists

Let the probability of recalling correctly a confusable item for position i be $p(C_i)$, and correctly recalling a non-confusable item be $p(N_i)$. Then the peaks of the sawteeth in alternating curves suggest that, for all i, the probability of failing to recall a confusable item is greater than the probability of failing to recall a non-confusable item, or:

$$p(\overline{C_i}) > p(\overline{N_i})$$

Now phonological similarity might conceivably impair recall at two stages: in retrieval of a response for position i and in cuing of that response by the items recalled in previous positions $1, \ldots, i-1$. The above inequality supports the former, but an effect of similarity at retrieval may also mask a smaller but significant effect of similarity at cuing. More importantly, the functional cue for response i may differ from the correct cue for the item in position i. In other words, previous responses may be in error and hence not match the predecessors of the correct item in the list. In this case, a conditional analysis of errors is required.

The probability of erring in recall of the *ith* confusable item can be rewritten as follows:

$$p(\overline{C_i}) = p(\overline{C_i} \cap N_{i-1}) + p(\overline{C_i} \cap \overline{N_{i-1}})$$
$$= p(\overline{C_i}/N_{i-1})\,[1 - p(\overline{N_{i-1}})] + p(\overline{C_i}/\overline{N_{i-1}})\,p(\overline{N_{i-1}})$$

Similarly, the recurrence relation for erring in recall of non-confusable items is:

$$p(\overline{N_i}) = p(\overline{N_i}/C_{i-1})\,[1 - p(\overline{C_{i-1}})] + p(\overline{N_i}/\overline{C_{i-1}})\,p(\overline{C_{i-1}})$$

Now there are many possible inequalities between components of the right-hand sides of these equations that are consistent with the inequality found empirically between the left-hand sides. For example, the Markovian chain across output positions can be broken with the simplifying assumptions that, for all i:

$$p(\overline{C_i}) = c \qquad p(N_i) = n$$

Then, dropping subscripts for clarity and substituting into the empirical inequality:

$$c = p(\overline{C}/N)\,(1 - n) + p(\overline{C}/\overline{N})\,n > p(\overline{N}/C)\,(1 - c) + p(\overline{N}/\overline{C})\,c = n$$

Putting $c = 4/9$ and $n = 1/3$, as rough approximations from Baddeley's results, there exists a solution such that:

$$p(\overline{N}/C) = \frac{1}{5} \qquad p(\overline{C}/N) = \frac{1}{3}$$

$$p(\overline{N}/\overline{C}) = \frac{1}{2} \qquad p(\overline{C}/\overline{N}) = \frac{2}{3}$$

This example shows it is possible that: (a) errors are more likely to follow other errors than correct responses, as would be expected; and (b) that errors following failure to recall a confusable item are less likely than errors following failure to recall a non-confusable item. If the failure to recall a confusable item reflects substitution of one confusable item for another, result (b) is readily explicable in chaining terms by a greater chance of erroneous confusable cues allowing recovery from an error. These possibilities cannot be detected with error position curves that estimate only unconditional probabilities.

Coincidence with Non-confusable Curve

The increased chance of recovering from errors following confusable items is enlightening when the second important finding of Baddeley's is considered, that the percentage errors in recalling non-confusable items in alternating lists is not significantly greater than recalling non-confusable items in non-confusable lists.

Let the probability of recalling the correct non-confusable item in any output position for alternating lists be $p(N_a)$ and the correct non-confusable item in any output position for the non-confusable lists be $p(N_p)$. Then Baddeley's error position curves suggest:

$$p(\overline{N_a}) = p(\overline{N_p})$$

Again, reexpressing the probability of not recalling these items in terms of failure to recall their immediate predecessor:

$$p(\overline{N_a}) = p(\overline{N_a}/C_a)\,[1 - p(\overline{C_a})] + p(\overline{N_a}/\overline{C_a})\,p(\overline{C_a})$$

$$p(\overline{N_p}) = p(\overline{N_p}/N_p)\,[1 - p(\overline{N_p})] + p(\overline{N_p}/\overline{N_p})\,p(\overline{N_p})$$

And letting:

$$p(\overline{C_a}) = c > n = p(\overline{N_a}) = p(\overline{N_p})$$

Then:

$$p(\overline{N_a}/C_a)\,(1 - c) + p(\overline{N_a}/\overline{C_a})\,c = p(\overline{N_p}/N_p)\,(1 - n) + p(\overline{N_p}/\overline{N_p})\,n$$

The above equality may be satisified simultaneously by both inequalities:

$$p(\overline{N_a}/C_a) > p(\overline{N_p}/N_p) \qquad 1$$

$$p(\overline{N_a}/\overline{C_a}) < p(\overline{N_p}/\overline{N_p}) \qquad 2$$

Continuing with the previous example, the following solutions obey all the empirical constraints, as well as observing inequalities 1 and 2:

$$p(\overline{C_a}) = \frac{4}{9} \qquad p(\overline{N_a}) = \frac{1}{3} \qquad p(\overline{N_p}) = \frac{1}{3}$$

$$p(\overline{N_a}/C_a) = \frac{1}{5} \qquad p(\overline{C_a}/N_a) = \frac{1}{3}$$

$$p(\overline{N_a}/\overline{C_a}) = \frac{1}{2} \qquad p(\overline{C_a}/\overline{N_a}) = \frac{2}{3}$$

$$p(\overline{N_p}/N_p) = \frac{1}{30} \qquad p(\overline{N_p}/\overline{N_p}) = \frac{14}{15}$$

Some hypothetical reports that fit these example probabilities are shown in Table 4. Given lists whose correct orders are represented by the sequence *X1 X2 X3 X4 X5 X6*, where *X* represents either a confusable (*C*) or

TABLE 4
Example Reports for Alternating AN and Non-confusable PN List Types

List Type AN[a]	*List Type PN*[a]
N1 C2 N3 C4 N5 C6	N1 N2 N3 N4 N5 N6
N1 C2 N3 C4 N5 C6	N1 N2 N3 N4 N5 N6
N1 C2 N3 C4 N5 C6	N1 N2 N3 N4 N5 N6
N1 C2 N3 C4 N5 C6	N1 N2 N3 N4 N5 N6
N1 $\overline{C6}$ N3 $\overline{C2}$ N5 $\overline{C4}$	N1 N2 N3 N4 N5 N6
N1 $\overline{C6}$ N3 $\overline{C2}$ N5 $\overline{C4}$	$\overline{N3}$ $\overline{N1}$ $\overline{N2}$ N4 N5 N6
$\overline{N5}$ C2 $\overline{N1}$ C4 $\overline{N3}$ C6	N1 N2 N3 $\overline{N5}$ $\overline{N6}$ $\overline{N4}$
$\overline{N3}$ $\overline{N1}$ $\overline{C2}$ $\overline{C6}$ $\overline{C4}$ $\overline{N5}$	$\overline{N3}$ $\overline{N1}$ $\overline{N2}$ $\overline{N6}$ $\overline{N4}$ $\overline{N5}$
$\overline{N3}$ $\overline{N1}$ $\overline{C2}$ $\overline{C6}$ $\overline{C4}$ $\overline{N5}$	$\overline{N3}$ $\overline{N1}$ $\overline{N2}$ $\overline{N6}$ $\overline{N4}$ $\overline{N5}$

[a] C = confusable item, N = non-confusable item, number refers to input position, and overlined responses are errors.

non-confusable (*N*) item, and the number refers to the item's input position, the errors for each output position in Table 4 are overlined. It can be verified easily that the distributions of errors in these reports are consistent with all the probabilities given above.

One caveat accompanies the above analysis. Breaking the Markov chain to simplify analysis effectively assumes both flat error position curves and that the cue used in chaining models consists of only the immediately preceding response. The former is in stark contradiction to experimental results, but its implications for the predictions of chaining models will be small. The latter is problematic for chaining theories in general, as can be seen by considering lists with repeated items. Thus many chaining models have the cue as some compound of a number of preceding responses. Analysis of these models along the same lines as above would be considerably more complicated. However, the implications will be the same: Any confusable item that forms part of the cue for the next response will decrease the ability of that cue to identify uniquely the next response.

How Properly to Test Chaining Models

As shown above, the sawtooth error position curves for alternating lists, whose troughs are coincident with non-confusable curves, do not unambiguously rule out chaining models. Rather, in order to fit these data, chaining models seem to have to make a number of simple predictions about conditional probabilities of errors following correct or incorrect, confusable or non-confusable responses. In particular, given veridical previous responses, the conditional probability of failing to recall a non-confusable item should be greater following a confusable item in alternating lists than following a non-confusable item in non-confusable lists, an effect of similarity at cuing (equation 1). However, though making the first error in recall of a non-confusable list may be less probable than in alternating lists, once it has been made, there may be a greater probability that further errors will be made (equation 2). In chaining terms, this is because an erroneous response is likely to cue its successor in the list (which will be a further error), and this possibility is greater when the erroneous cue is not confusable. In other words, relative errors should be more frequent in recall of non-confusable lists than alternating lists (and, in turn, relative errors should be more frequent in recall of alternating lists than confusable lists).

One reason that these predictions have been overlooked in the past has been due to the failure to identify the functional cue that is involved in the chaining. Chaining models assume that cues are the previous responses in a report, which may be incorrect and so cannot simply be assumed to be the previous items in the list. The functional cue cannot be identified from error position curves alone. Rather, what are needed are conditional error position curves, which give the probability of making an error given that all previous responses are correct.

Again, chaining models do not exclude the possibility that phonological similarity acts in retrieval, in selection of a response, as well as in cuing a set of likely responses. The sawteeth suggest that phonological similarity does act at retrieval, and, furthermore, the above analysis suggests that the peaks of the sawteeth should reflect confusable items being substituted for one another. Validating this selective substitution is another reason for attempting a replication of Baddeley's experiment and looking more closely at the nature of errors.