Chapter 2: Chaining Theory

An experiment testing Chaining Models

This chapter describes an experiment involving recall of lists of alternating confusable and nonconfusable items. The results, together with those in Henson et al. (1996), are troublesome for chaining models of serial recall. The chapter also includes a detailed analysis of transpositions in serial recall, which is used to test the alternative model in Chapter 5.

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, P,* is harder to recall than the order of items that are pronounced differently, such as *C, F, J* (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 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).

Wickelgren (1965b) offered an explanation for the phonological similarity effect in terms of a simple chaining model, where items are stored by pairwise associations between their constituent phonemes. Assuming each phoneme has a single (type) representation in memory, repeated phonemes, such as the vowel */i:/* in the list *B, D, P,* are associated with more than one successor (i.e., */d/* and */p/*). Such lists are therefore formally equivalent to lists with repeated items, and the phonological similarity effect occurs for the same reason as associative intrusions (Chapter 1; Wickelgren, 1966). That is, phonological similarity acts on the cuing of items, because repeated phonemes are ambiguous cues for their successors.

A similar prediction would appear true of other chaining models. The most obvious way to model phonologically similar items in TODAM (Murdock, 1983) and recurrent networks (e.g., Jordon, 1986) would be to assume overlapping (nonorthogonal) vector representations. This would also produce an effect of similarity on cuing. Indeed, a general property of such distributed, associative memories is that "...errors are more likely when discriminations must be made between similar states..." (Jordan, 1986, p. 37). The exact predictions of a compound chaining model that chains along phonological representations are shown analytically in Henson (1994).

Baddeley (1968; Experiment V) tested whether phonological similarity affects the cuing of items, as suggested by chaining models, or whether it affects the retrieval of items. He used immediate serial recall of lists of six items, where the items were drawn from a set of letters pronounced similarly (the *confusable items*) and a set of letters pronounced differently (the *nonconfusable items*). With lists in which confusable and nonconfusable items alternated, error position curves revealed a "sawtooth" shape, where the peaks of the sawteeth represented errors in recall of confusable items, and the troughs represented fewer errors in recall of adjacent nonconfusable items (e.g., Figure 2-1). The sawteeth for these *alternating lists* were confined within more conventionally bowed curves for two *pure lists*: the *confusable lists*, which contained only confusable items, and the *nonconfusable lists*, which contained only nonconfusable items. While the peaks of the sawteeth lay below the curve for confusable lists, the troughs were virtually coincident with the curve for nonconfusable lists.

Baddeley argued that the fact that most errors in recall of alternating lists occurred for confusable items, rather than the nonconfusable items that followed them, favoured the idea of phonological similarity acting on retrieval rather than on cuing. Indeed, the fact that the confusable items in alternating lists had little to no effect on recall of the nonconfusable items, when compared with those in nonconfusable lists, suggested that there is no effect of phonological similarity on cuing.

Disregarding chaining models on the basis of these results is premature however. Sawteeth on their own are certainly insufficient. This is because chaining models could predict an effect of similarity on retrieval as well as on cuing (e.g., at the deblurring stage of TODAM; Lewandowsky & Li, 1994). Sawteeth could then result if the effect of phonological similarity is simply greater on retrieval than on cuing. The apparent coincidence of alternating and nonconfusable curves, for recall of nonconfusable items, is harder to reconcile with chaining models. However, this coincidence was not found in Experiment VI of the same paper, which used auditorily presented words. Moreover, a more sophisticated, probabilistic analysis shows the combined effect of phonological similarity on cuing and on retrieval can, in principle, reconcile chaining models with Baddeley's data (Henson, 1994; Henson et al., 1996).

Experiment 1

The first aim of Experiment 1 was to replicate Baddeley's results with a more powerful design geared towards detecting an effect of phonological similarity on cuing. The most important comparison was between recall of nonconfusable items in alternating lists (where they were preceded by a confusable item) and recall of nonconfusable items in nonconfusable lists (where they were preceded by another nonconfusable item). An impairment in recall of nonconfusable items when their predecessors were phonologically similar to other list items would constitute evidence for chaining models.

A second aim was to conduct a more thorough analysis of subjects' responses. Though Baddeley reported errors by position, he did not examine the actual types of error, such as whether the errors were omissions or substitutions. Such analysis addresses further theoretical questions. For example, some theories suggest that similar representations degrade faster than 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 have reflected a greater incidence of confusable items being omitted, or being substituted for a random guess. However, if phonological similarity acts through response competition during retrieval, then the majority of these errors should be confusions; that is, one confusable item being substituted for another (e.g., Bjork & Healy, 1974; Conrad, 1965). As shown in Henson et al. (1996), this type of substitution is important if chaining theories are to be reconciled with Baddeley's data.

One modification in design of the present experiment was to generate the lists from a small experimental vocabulary, and to block the conditions separately, rather than intersperse them randomly as in Baddeley's experiment. This ensured that all lists in a block contained the same six items (conforming to the "order only" condition of Healy, 1974). With such a design, subjects know in advance which particular items will be presented, and need only concentrate on the order in which they occur. Consequently, minimal numbers of intrusions and omissions were expected, making transpositions the most likely errors. This allowed the simplifying assumption that reports were permutations of list items, and hence determination of the chance probabilities of certain responses.

A further interest was the distribution of associate errors (Chapter 1). Given an effect of phonological similarity on cuing, phonological chaining models predict that associates will be more frequent for nonconfusable than confusable lists. This is because a nonconfusable item is more likely to cue its successor in the list than is a confusable item (which partially cues other items; Henson, 1994). More generally, any closed-loop chaining model predicts that the frequency of associates should exceed that expected by chance, irrespective of phonological similarity. This is because the erroneous item, even if only part of the cue for the next response, will still increase the probability of recalling its successor, rather than the correct successor. These constituted two more specific tests of chaining theory.

A final modification in design was that subjects in the present experiment were encouraged to group the six items into two groups of three. Baddeley did not give such instruction to his subjects. However, grouping strategies are often brought to bear on the most simple of span tasks, and can have dramatic effects on the pattern of errors (Chapter 3). Particular advantage is conveyed to recall of the first and last items in a group, revealed as primacy and recency effects within each group. Indeed, a suggestion of such spontaneous grouping is apparent in Baddeley's error position curves, particularly for confusable lists. The concern was that grouping strategies might interact with the structure of alternating lists. For example, a choice of grouping in twos rather than threes may affect the nature of errors made in recall of alternating lists. Thus the explicit instruction to group in threes in the present experiment was to encourage a single, consistent grouping strategy across subjects.

In summary, the aims of the experiment were: 1) to reproduce and make explicit tests of Baddeley's findings, specifically the sawtooth error position curves for alternating lists; and 2) to conduct a more thorough analysis of errors.

Method

Subjects

Forty-eight subjects from the APU Subject Panel were tested, of whom seventeen were male and thirty-one were female. Their mean age was twenty-seven years.

Materials

Stimuli were lists of six, single-syllable consonants, generated from a vocabulary of twelve. The letters were classified according to their confusability; that is, whether they were phonologically similar to any other letters in the vocabulary. The six confusable letters shared a common rhyme when pronounced: *B, D, G, P, T, V;* the six nonconfusable letters possessed unique rhymes: *H, K, M, Q, R, Y.*

The two pure list types were the confusable lists, containing all six confusable letters, and the nonconfusable lists, containing all six nonconfusable letters (conditions PC and PN respectively). Two alternating list types (A1 and A2) were identified according to the two mutually exclusive sets of three confusable and three nonconfusable letters in the vocabulary (Table 2-1). These lists comprised the two alternating conditions, according to whether the alternation began with a confusable or a nonconfusable item in the first position (conditions AC and AN respectively). Conditions AC and AN were nested inside list types A1 and A2, such that a block of A1 or A2 lists contained six lists of condition AC and six of condition AN. With the randomised order of lists within blocks, this nesting was to reduce the chance of subjects' detecting a pattern of confusable-nonconfusable alternation.

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

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 six letters that would appear in the following 12 lists were presented in a circle, in order to familiarise 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 randomised and the block order was fully counter-balanced across subjects.

The experiment was run on an IBM PC, with the capitalised letters appearing in the

Condition	List Type	List Structure	Letter Set (Example List)	Number of Lists
PC	PC	CCCCCC	BDGPTV	12
PN	PN	NNNNNN	HKMQRY	12
AC	A1	CNCNCN	DQTMPK	6
	A2	CNCNCN	BHGYVR	6
AN	A1	NCNCNC	QDMTKP	6
	A2	NCNCNC	HBYGRV	6

Table 2-1: Composition of lists in Experiment 1.
(C=confusable item, N=nonconfusable item).

centre of a monochrome VDU, each letter approximately half an inch high and replacing the previous one. Presentation rate was two letters per second (400-ms on; 100-ms off). Subjects were instructed to read the letters in silence, and immediate recall was prompted after the last item disappeared, with letters written left to right across a row of six boxes provided on a response sheet. A minimum of 10 seconds was required between trials, after which subjects pressed a key to start the next trial. A short break of a minute occurred between blocks.

Subjects were instructed to write down answers immediately and, if unsure, told to "write the first letter that comes to mind". If no letter came to mind, 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

In brief, the results replicated those of Experiment V in Baddeley (1968), though there was a confounding effect of the predictability of lists (Henson et al., 1996). Nevertheless, there were significant sawteeth in alternating conditions, in addition to normal primacy and recency. Closer analysis showed that the peaks of these sawteeth reflected confusable items transposing with one another, and that such confusions were sensitive to transposition distance. Most importantly, there was little evidence for an effect of similarity on cuing, even taking into account predictability, while there was clear evidence for an effect on retrieval. This was apparent in both error position curves and more detailed analyses of associate errors.

Overall Performance

Approximately 20% of PC lists, 58% of PN lists, 55% of AC lists and 51% of AN lists were recalled correctly. Omissions comprised approximately 5% of errors, while intrusions amounted to only 3%. The rarity of such errors reflected the small experimental vocabulary, and probably accounts for the higher level of performance than in Baddeley's experiment.

Predictability

Error position curves (upper panel of Figure 2-1) replicated the main features of Baddeley's. The effect of phonological similarity extended over all positions in the pure confusable lists, but just the positions of confusable items in the alternating lists. Importantly, there was no evidence of more errors for nonconfusable items in alternating lists than in pure nonconfusable lists. In fact, nonconfusable items were recalled slightly better in alternating lists (i.e., the sawteeth straddled the nonconfusable curve, rather than sitting on top of it).

As reported in Henson et al. (1996), closer inspection of the stimuli suggested a reason for this: The letters in different list types differed in their *predictability* (e.g., how often the letters cooccur in the English language; Baddeley, Conrad & Hull, 1965). Letters in the A2 lists were especially predictable. This might explain why performance in alternating conditions AC and AN was slightly better than expected, compared to condition PN.

The counterbalanced design of lists meant that predictability should not affect tests within conditions. However, predictability did potentially confound tests across conditions. Though the larger experimental vocabulary in Baddeley's experiments made such a confound less likely, a similar caution should apply to his results also. Without equating predictability across conditions, one cannot be sure that performance on nonconfusable items in alternating lists was truly unaffected by the presence of confusable items.

Two further experiments in Henson et al. (1996) controlled for the predictability of letters (as did all subsequent experiments in the present thesis). The approach taken here, when comparing across conditions, was to remove the A2 lists from analysis, so that the AC

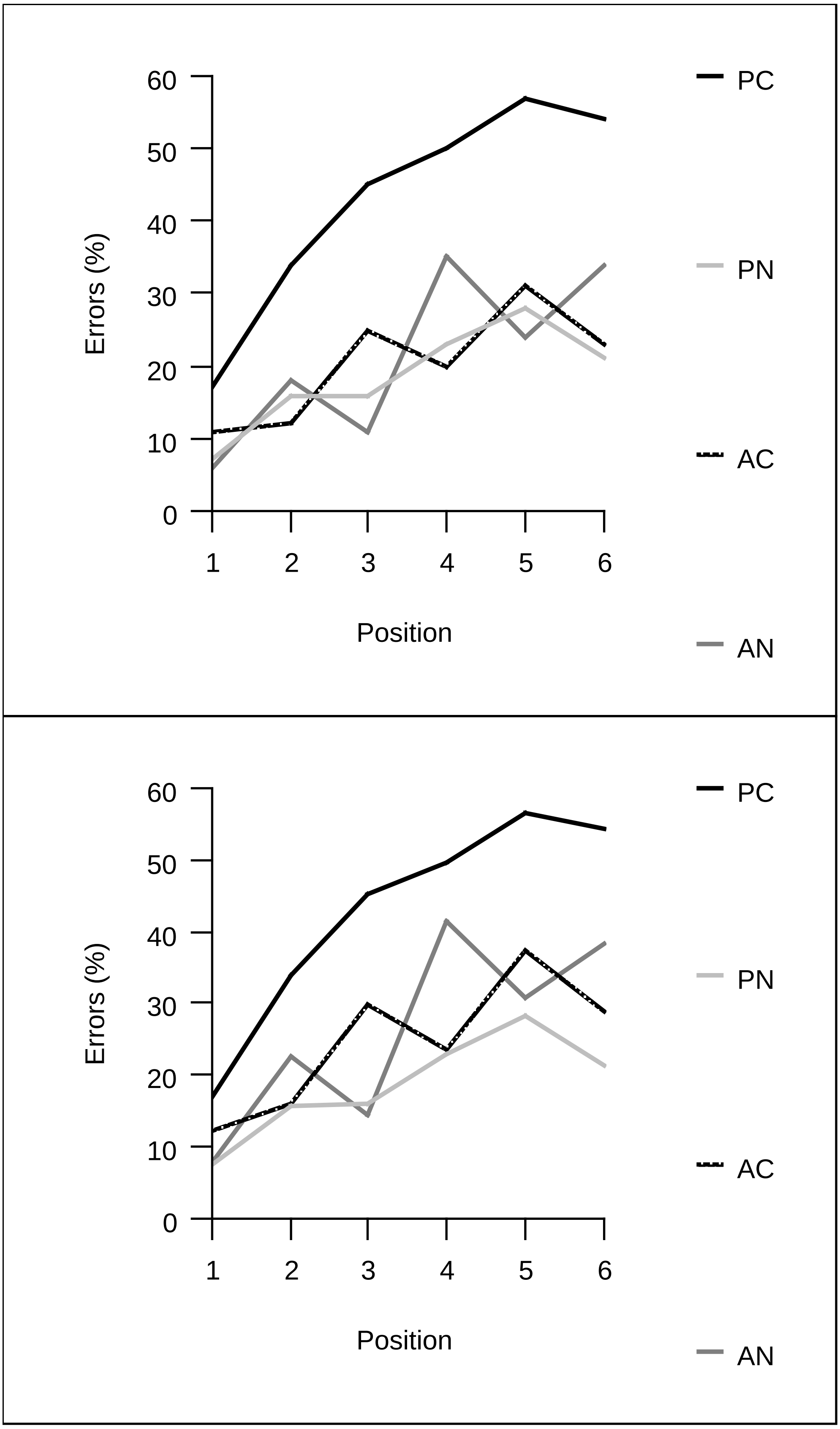

Figure 2-1: Errors by position including A2 lists (upper panel) and excluding A2 lists (lower panel) in Experiment 1.

and AN conditions were calculated from the A1 list type only (lower panel of Figure 2-1). The predictability of letters in A1 lists was less than that in PN and PC lists (Henson et al., 1996), which will tend to increase errors in alternating lists compared to the pure lists. A failure to find significantly more errors for nonconfusable items in conditions AC and AN than PN can not therefore be attributed to different predictabilities.

Comparisons within Conditions

In order to test Baddeley's findings statistically, planned comparisons were performed on the log-odds of an error (Chapter 1) across each position in the upper panel of Figure 2-1 (i.e., including A2 lists) in a separate ANOVA for each condition.

Two 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 last-item recency (the error score on Position 6 versus Position 5). Both confusable and nonconfusable curves showed significant primacy, $F(1,235)>18.00$, $MSE<0.52$, $p<.001$, but only the nonconfusable curve showed significant recency, $F(1,235)=8.18$, $MSE=0.50$, $p<.01$ ($F<1$ for the confusable curve).

Three comparisons for alternating lists AC and AN tested the significance of the sawtooth shape (the error score on confusable positions compared to adjacent nonconfusable positions). A fourth contrast looked for an effect of primacy over the first four positions (as defined above). For both alternating curves, errors were significantly more common on confusable positions than adjacent nonconfusable positions, $F(1,235)>4.62$, $MSE<0.57$, $p<.05$ in all cases, except between the first two positions of condition AC, $F<1$. The latter probably reflected the opposing effect of primacy, which was significant in both AC and AN conditions, $F(1,235)>42.25$, $p<.001$.

Comparisons between Conditions

To test the predictions of phonological chaining models, the weighted log-odds of an error on nonconfusable positions in alternating lists was compared to that on nonconfusable positions in nonconfusable lists. Including A2 lists, there was no significant difference on any of the six positions, $Z(48)<0.39$, $p>.70$. This may have reflected the less predictable nature of the A2 lists. Even excluding A2 lists however, there was still no greater probability of an error in alternating lists for any position, $Z(48)<1.81$, $p>.07$, except the last, $Z(48)=3.16$, *family-wise* $p<.01$ (using Holm's correction for the multiple comparisons). Thus, when examining the troughs of the sawteeth in Figure 2-1, there was only evidence for an effect of similarity on cuing for one of the six positions, providing A2 lists were excluded.

In addition, the weighted log-odds of an error on confusable positions in alternating lists was compared to that on corresponding nonconfusable positions in nonconfusable lists. Both including and excluding A2 lists, there was a significantly greater probability of errors on all six confusable positions, $Z(48)>3.21$, *family-wise* $p<.01$. Thus, when examining the peaks of the sawteeth in Figure 2-1, there was evidence for an effect of similarity on retrieval for all positions, whether or not A2 lists were excluded.

Finally, the weighted log-odds that adjacent transpositions were associates was calculated for confusable and nonconfusable lists, for the 33 subjects who made at least one pair of adjacent transpositions in both conditions. There was no evidence for a greater probability of associates in nonconfusable lists ($M=.23$, $SD=.22$) than confusable lists ($M=.22$, $SD=.15$), $Z(33)=0.08$, $p=.94$; another failure to find any effect of similarity on cuing. Nor did these probabilities differ significantly from a chance probability of *.20* (assuming the second error could be one of five list items), for either nonconfusable lists, $Z(33)=0.23$, $p=.82$, or confusable lists, $Z(33)=0.25$, $p=.80$, contrary to closed-loop chaining models.[1]

Transpositions

Transposition gradients were also calculated for each condition, collapsing across subjects (Figure 2-2; the six bars for each output position represent the percentage of responses from each input position, from left to right, the tall bars being correct responses). For PC and PN lists, transpositions decreased monotonically with increasing distance between input and output position. This monotonic decrease was remarkably lawful: The rank ordering of transpositions for each output position would be expected only 1 in 120 times if subjects guessed list items at random. The only exception to this monotonic decrease was transpositions from the first to the last position in PN lists; further inspection revealed that these were mainly repetitions (Chapter 4). The transposition gradients for the AN lists did not

1. Even if the mean probability were *.30*, the null hypothesis could not be rejected at the *.05* level (*power=99%*).

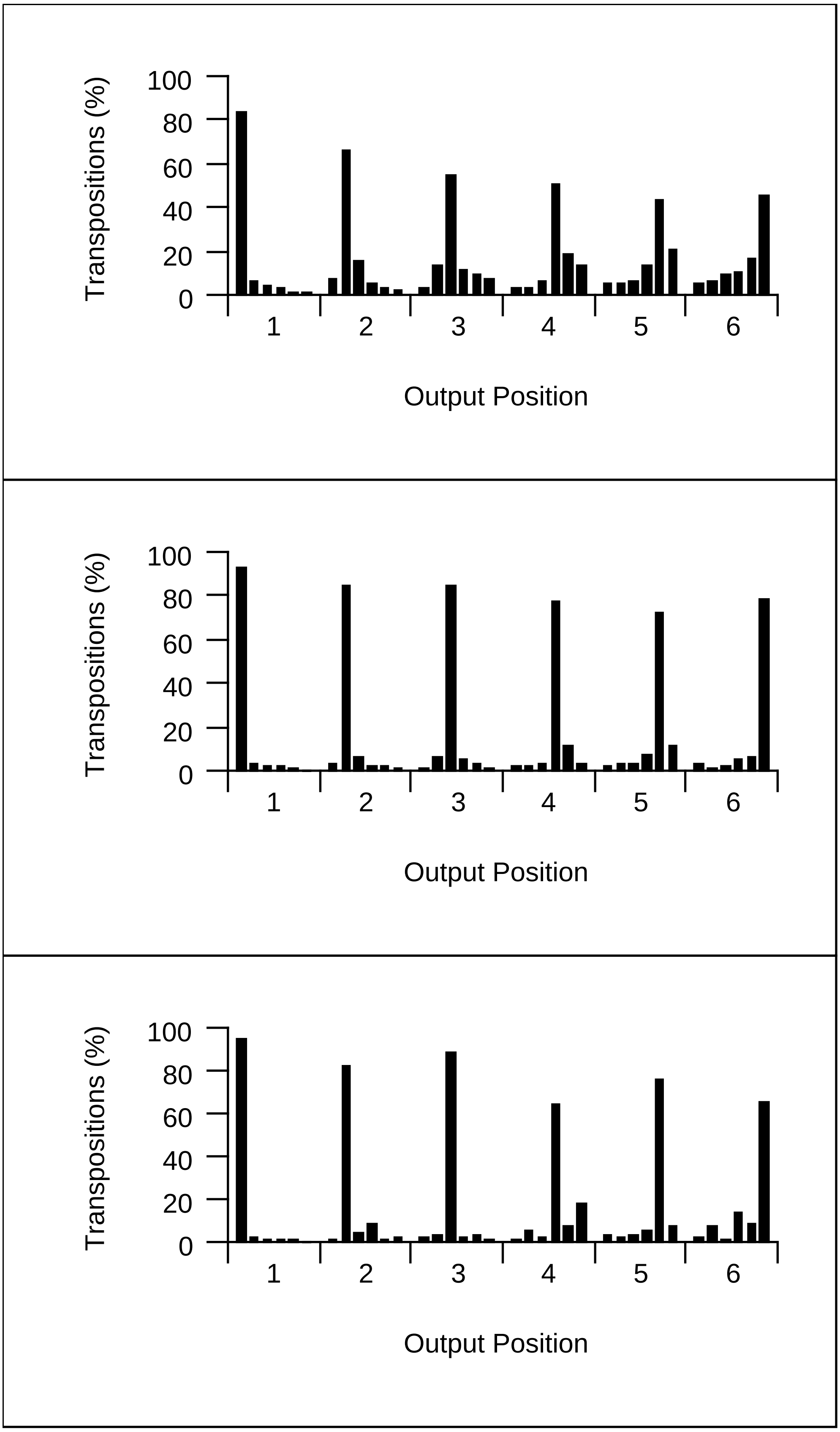

Figure 2-2: Transposition gradients for conditions PC (upper panel), PN (middle panel) and AN (lower panel) in Experiment 1.

decrease monotonically, but were a function of the phonological similarity between the correct and transposed item. Thus transposition gradients for output positions that corresponded to input positions of confusable items (Positions 2, 4 and 6) were punctuated by peaks for input positions of other confusable items. The same pattern arose for AC lists. Because the majority of reports were in effect permutations of list items, given that most errors were nonrepeated transpositions, the sawteeth shape of error position curves logically requires that the majority of transpositions in alternating lists were confusable items transposing with one another.

When transpositions were replotted against transposition distance, the gradients for confusable and nonconfusable lists were not parallel: The gradients were steeper for confusable curves (Figure 2-3), a finding confirmed in Chapter 4. In other words, the effect of phonological similarity was not additive, implying that phonological confusions do not arise independently of position (Chapter 5).

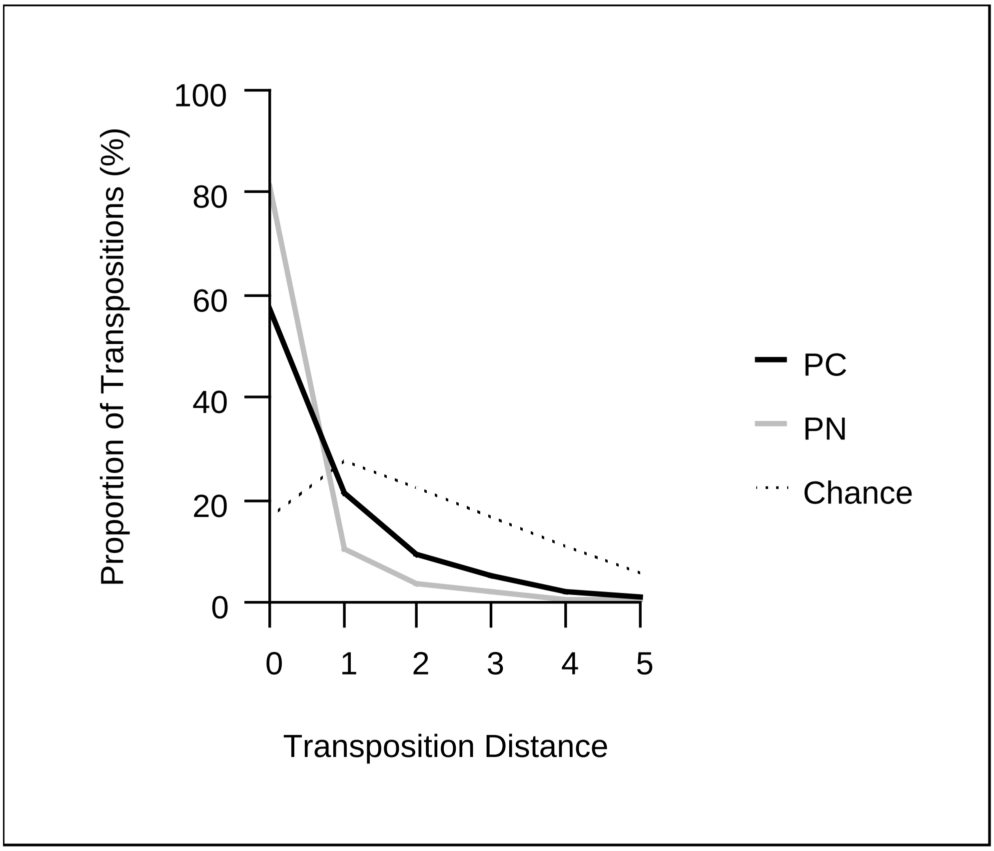

Figure 2-3: Proportion of transpositions (including correct responses) by transposition distance, together with chance levels, in Experiment 1.

Discussion

Results from the present experiment, and others in Henson et al. (1996), consistently fail to provide any support for chaining theory. In fact, they are difficult to reconcile with any current chaining model.

The prominent sawteeth in the error position curves for alternating lists reflect significantly more errors in recall of confusable than nonconfusable items. Though chaining models can be constructed that are compatible with this sawtooth pattern, they remain unable to explain the fact that, whether unconditional (or conditional, Henson et al., 1996) error probabilities are examined, the presence of confusable items in a list most often has no detectable effect on the probability of recalling following nonconfusable items. In the present experiment for example, only one nonconfusable position in six showed significantly more errors in alternating curves that nonconfusable curves, and this could owe to the less predictable nature of the alternating lists (with A2 lists removed). Though one can never be certain whether the failure to find a significant difference on the other five positions reflected a lack of statistical power, the same experiment was powerful enough to show significantly more errors on all six confusable positions in alternating curves. These findings meet Frick’s (1995) criteria for accepting the null hypothesis at least (though see Chapter 4).

Nevertheless, a number of questions might be asked of the present results. Firstly, there is the question of whether the phonological confusions arose during encoding rather than retrieval. Confusions during encoding would seem unlikely with visual presentation however; none was observed when subjects read items aloud during presentation (Henson et al., 1996, Experiment 3). Even with auditory presentation, an encoding account seems insufficient (Baddeley, 1968, Experiment IV). Secondly, there is the question of strategic effects: Subjects might treat confusable and nonconfusable items differently in alternating lists (e.g., by streaming them, particularly with auditory presentation; Jones, 1992). This also seems unlikely, given that not one subject in the present experiment detected the alternating pattern in A1 and A2 lists. Finally, there is the question of generality: The present results did not hold perfectly for subjects with low memory spans (Henson et al., 1996), who showed less well-defined sawteeth. This may reflect subjects abandoning phonological coding when they approach the limit of their memory (Salame & Baddeley, 1986). More likely, the differences reflect "knock-on" effects when recall becomes difficult and subjects "give up" (Chapter 4). In any case, conditional probabilities of first errors (Henson et al., 1996), which remove knock-on effects of prior errors, showed the same pattern as the present results.

These questions aside, there are two aspects of the present data that are troublesome for chaining theory: 1) recall of nonconfusable items was little affected by whether the previous item was confusable, and 2) recall of nonconfusable items was little affected by whether or not the previous item was confused. In other words, there was little evidence for an effect of similarity on cuing, or an effect of errors on cuing.

The lack of any effect of similarity on cuing was reinforced by the probability of associative errors, which did not depend on whether or not the previous item was confusable. More generally, the present results suggest that confusable items have little effect on any surrounding nonconfusable items. A similar conclusion was reached by Bjork and Healy (1974): "...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).

The lack of an effect of similarity on cuing is troublesome for models that chain along phonological representations, such as that of Wickelgren (1965b). It appears troublesome for more recent models too, as soon as they adopt phonological representations (e.g., Murdock, 1983; Jordan, 1986). In particular, the inability of TODAM to simulate the Baddeley (1968) data was confirmed by Baddeley, Papagno and Norris (1991). The problems with chaining along phonological representations were also confirmed by Burgess and Hitch (1992), who obtained their best fits when associations between phonemes were minimised (in favour of position-item associations). The fact that present results hold when stimuli are vocalised (Henson et al., 1996, Experiment 3) is also problematic for theories that restrict chaining to the auditory modality (e.g., Drewnowski, 1980a; Penney, 1989).

One might argue that models that chain along nonphonological representations (e.g., context-sensitive tokens, Wickelgren, 1969) would not have to predict an effect of similarity on cuing. For example, TODAM might retain its normal assumption of random vector representations of items, and model phonological similarity as affecting only retrieval, during the subsequent "deblurring" of the results of chaining (Lewandowsky & Li, 1994). Another example is Richman and Simon's (1994) EPAM model. This model chains along unitised representations, or *chunks*, and locates phonological similarity in the retrieval of chunks. However, both these models face problems with the second aspect of the present data. As closed-loop chaining models, they still predict more errors on nonconfusable positions in alternating lists than in nonconfusable lists, because previous responses are more often in error in alternating lists. The additional errors on confusable items in alternating lists mean that the cue for the following nonconfusable item is correct less often than in nonconfusable lists. This prediction for an effect of errors on cuing was supported by neither error position curves, nor the incidence of associate errors, which were not significantly above chance.

The only type of chaining model consistent with present data would be an open-loop, nonphonological model (i.e., one that chained along nonphonological representations independently of feedback of previous responses). Such a specific model loses some of the intuitive appeal of chaining theory (e.g., that each response becomes the stimulus for the next). Moreover, given that there does not appear to be any data necessitating item-item chaining (Chapters 1, 4), and yet there is data necessitating positional information in short-term memory (Chapter 3), such a model does not seem worth pursuing.

The present results have in fact proved difficult for many models of serial recall, whether or not they employ chaining. For example, Burgess and Hitch (1992) stated in their abstract: "the model was unable to simulate human memory for sequences containing mixtures of phonemically similar and dissimilar items". This, together with comments in Henson et al. (1996), led to revision of the model (Burgess & Hitch, 1996a, 1996b). As such, the sawtooth shape of alternating curves is a bench-mark test for models of immediate serial recall (Page & Norris, 1996b). Chapter 5 presents a new model that can simulate memory for mixtures of phonemically similar and dissimilar items, and which passes this test.

Finally, the present experiment produced transposition gradients that replicate previous findings that items are more likely to transpose to nearby positions than positions far apart, the *locality constraint* (Chapter 4). The additional transpositions between confusable items suggest phonological similarity acts on retrieval, rather than via passive decay or interference during storage (Posner & Konick, 1966). That such confusions also respect the locality constraint is a new finding, which turns out to be important for modelling phonological similarity in short-term memory (Chapter 5).

Chapter Summary

This chapter examined the first of the three theories of serial order in Chapter 1: chaining theory. The fact that phonologically confusable items had little detectable effect on recall of surrounding nonconfusable items, either through an effect of similarity on cuing, or through an effect of errors on cuing, is difficult for current chaining models to explain. Though one might construct a very specific chaining model to fit the present data, the onus would be on the modeller to demonstrate additional evidence for such specific assumptions. Moreover, given that the next chapter demonstrates evidence for an alternative theory of serial order, there seems little point in pursuing a chaining theory of short-term memory for serial order.