Psychological Review
1998, Vol. 105, No. 4, 761–781

0033-295X/98/$3.00

The Primacy Model: A New Model of Immediate Serial Recall

Michael P. A. Page and Dennis Norris
Medical Research Council Cognition and Brain Sciences Unit

A new model of immediate serial recall is presented: the primacy model. The primacy model stores order information by means of the assumption that the strength of activation of successive list items decreases across list position to form a primacy gradient. Ordered recall is supported by a repeated cycle of operations involving a noisy choice of the most active item followed by suppression of the chosen item. Word-length and list-length effects are attributed to a decay process that occurs both during input, when effective rehearsal is prevented, and during output. The phonological similarity effect is attributed to a second stage of processing at which phonological confusions occur. The primacy model produces accurate simulations of the effects of word length, list length, and phonological similarity.

Serial recall from short-term memory is one of the most intensively studied tasks in cognitive psychology. To a large degree, interest in serial recall stems from a conviction that it involves a system whose operation underlies performance in a great variety of cognitive tasks. However, despite the theoretical importance of the topic, no detailed mechanism has yet been proposed that can give an integrated explanation of the central body of empirical data on immediate serial recall (ISR). Although researchers using theories like the *working memory* model (Baddeley, 1986; Baddeley & Hitch, 1974) have developed qualitative accounts of a range of phenomena, such as the effects of phonological similarity and word length, theorists who have attempted to develop a more quantitative account of the processes underlying serial recall have typically been more restricted in the range of ISR data that they seek to explain (e.g., Drewnowski, 1980; Estes, 1972; Lee & Estes, 1981; Lewandowsky & Murdock, 1989; Richman & Simon, 1994; Shiffrin & Cook, 1978).

The modeling of immediate serial recall has been dominated by approaches that emphasize two possible underlying mechanisms: recall by means of position–item associations and recall by means of chains of item–item associations.[1] This dichotomy was one that was addressed extensively within a related (but distinct) literature on serial-list learning (e.g., Ebenholtz, 1963; Jensen & Rohwer, 1965; Slamecka, 1967; Young, 1961, 1962, 1968), though without a clear resolution in favor of either mechanism. For immediate serial recall, the position–chaining distinction has been addressed most explicitly by Burgess and Hitch (1992), whose simulations led them to conclude "that simulations with little or no chaining came closest to reproducing human behavior, particularly in relation to order errors and the shape of the serial position curve" (p. 456). Even when using purely positional information, however, their model fails to reproduce important aspects of the serial-recall data, such as pronounced primacy and recency effects and the detailed pattern of errors underlying the phonological similarity effect (Baddeley, 1968; Henson, Norris, Page, & Baddeley, 1996; but see Burgess, 1995). Here we introduce a simple model that explains all these effects and others and that relies neither on item–item chaining nor on positional information but on a more direct coding of item order.

Scope of the Model

The first, and most obvious, target for any model of serial recall is to reproduce the basic serial-position curve.[2] The serial-position curve has a characteristic shape: a *primacy* portion, in which errors increase with serial position, and a *recency* portion (frequently confined to the last item in the case of visually presented lists), in which errors decrease with list position. It is very important to note that this *recency effect* is qualitatively different from that found in free-recall experiments, in which recall order is unconstrained and the items presented last are typically recalled first. In serial recall, the advantage for the final item occurs in spite of the fact that it is recalled last.

Serial recall gives rise to qualitatively different types of error:

Michael P. A. Page and Dennis Norris, Medical Research Council Cognition and Brain Sciences Unit, Cambridge, United Kingdom.

This research was supported by Joint Councils Initiative Grant E304/187.

Parts of this work were reported at the International Conference on Working Memory, Cambridge, United Kingdom, July 1994. We thank Alan Baddeley and Rik Henson for their invaluable assistance throughout the course of the work described here. We also thank Elizabeth Maylor, Nelson Cowan, and Alice Healy for their helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Michael P. A. Page, Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom. Electronic mail may be sent to mike.page@mrc-apu.cam.ac.uk.

[1] Note that chaining accounts require at least one position–item association, namely, that with the first item, to initiate recall.

[2] Note that we use the term *serial recall* throughout to refer to tasks conducted to ensure that recall proceeds in strict forward order, with no backtracking, scored such that an item is considered correct only if recalled in the correct position. Participants are able to indicate an omission either by saying "blank" during verbal recall or by drawing a line during written recall.

substitutions, whereby an item is replaced with another item in recall, or *omissions,* whereby no item is recalled in a given position. Substitutions can be further classified into *transpositions,* whereby an item is replaced with another item from the same list, and *intrusions,* whereby an item is replaced with an item from outside the target list. Repeats in recall, where there was none in the stimulus, can fall into either category, though a repeated intrusion is extremely unusual. The distribution of such errors in recall can tell us a great deal about the mechanisms underlying recall. Our aim, therefore, is not only to reproduce the serial-position curve but also to account for the detailed pattern of errors underlying it.

The intended scope of our model is best discussed with reference to the working memory framework established by Baddeley and Hitch (1974) and further developed by Baddeley (e.g., Baddeley, 1986). In Baddeley's model, immediate serial recall is mediated by the so-called *phonological loop,* comprising a phonological store and an articulatory control process capable of supporting subvocal rehearsal. Consideration of this model has motivated a series of experiments whose results place strong constraints on models of short-term memory, in particular models of immediate serial recall. A number of more specific observations have been made, each of which should be explained by a detailed model of serial recall.

First, performance in recalling a list composed of items that are phonologically confusable, such as the rhyming letter names *B, C, D,* and *G,* is worse than that for low-confusability stimuli, such as the letter names *H, Q, R,* and *Y* (Baddeley, 1966, 1968; Conrad, 1964; Conrad & Hull, 1964; Wickelgren, 1965a, 1965b). This *phonological similarity effect* is found with both auditorily and visually presented lists, but for visually presented lists the effect disappears with articulatory suppression during input (Murray, 1967, 1968; Peterson & Johnson, 1971). It is hypothesized that articulatory suppression interferes with the phonological recoding of visually presented stimuli.

Second, serial-recall performance is affected by the length of the words making up the lists. This effect is most clearly seen when comparing the recall of lists made up of one-syllable words with the recall of those comprising five-syllable words (Baddeley, Thomson, & Buchanan, 1975; Hulme, Maughan, & Brown, 1991), but Baddeley et al. also tested recall with words matched both for number of syllables and number of phonemes to show that articulation rate is the important determinant of performance. The typical finding (e.g., Hulme et al., 1991) is that participants' *span* (defined most often as the list length for which participants recall 50% of lists correctly) is linearly related to the rate at which they can articulate the stimulus items. Some aspects of this conclusion have been questioned recently (Caplan, Rochon, & Waters, 1992; Caplan & Waters, 1994; but see Baddeley & Andrade, 1994). Nonetheless, Baddeley's (1986) general account has been successful in explaining cross-linguistic differences in digit span (N. C. Ellis & Hennelly, 1980; Hoosain, 1987; Naveh-Benjamin & Ayres, 1986; Stigler, Lee, & Stevenson, 1986).

Third, recall is worse for longer lists. Performance, measured as the percentage of lists correctly recalled, when plotted against list length results in an inverse S curve (e.g., Drewnowski & Murdock, 1980).

Fourth, recall performance in response to an auditorily presented list exceeds that in response to the same list presented visually, the advantage manifesting itself toward the end of the list. The way in which we simulate this *modality effect* illustrates our belief that the benefits of auditory presentation reflect a component of memory additional to memory for order.

Before describing our model, we note that Baddeley's (1986) conception of the phonological loop is intimately tied to the immediate serial-recall task. There is a great deal of experimental evidence that indicates both the phonological nature of the memory system underlying performance in this task and the fact that representations in this memory system are extremely labile, being rendered useless by filled delays of as little as 2–3 s (e.g., Baddeley & Scott, 1971; Bjork & Healy, 1974; Conrad, 1967; Estes, 1973; Houston, 1965; Muter, 1980; Peterson & Johnson, 1971; Tehan & Humphreys, 1995). Thus, in presenting a model of immediate serial recall, we see ourselves as developing a computational model of the phonological loop, to complement Baddeley's verbal description. For this reason our model, like Baddeley's, should not be thought of as being applicable to tasks involving serial recall after significant delays.

The Primacy Model

General Approach

Rather than deriving order from a chain of interitem associations or from associations with positions (entities that are in some sense preordered), the primacy model stores order in terms of the relative activation levels of list items. In the description that follows, we assume that a given familiar stimulus item has, in connectionist terms, a *localist* representation, that is, a single node that corresponds to that stimulus. We assume that the order of items within a list can be represented by a pattern of activations across the corresponding nodes. More specifically, we assume that the order of items in the list "1234" can be represented by a pattern of activation such that $x_1 > x_2 > x_3 > x_4 > 0$, where x_i refers to the activation of the node representing the stimulus i. Thus, the rank ordering of the node activations corresponds to the temporal ordering of the stimuli. We term such a pattern of activation a *primacy gradient* of activations. Representing order in this way is not a new idea, having been described by Grossberg (1977, 1978) and subsequently employed by, for instance, Cohen and Grossberg (1987), Nigrin (1993), and Page (1993, 1994). These studies have received some neurophysiological support from investigations into long-term potentiation in response to sequences (Granger, Myers, Whelpley, & Lynch, 1993; Granger, Whitson, Larson, & Lynch, 1994; Lynch & Granger, 1992). Nevertheless, such a representation has not previously been employed in a detailed model of immediate serial recall.

In the earlier work cited above, and in some of our own recent work (Page & Norris, 1998; see the Appendix), a primacy gradient of activations is generated directly by presentation of the stimulus list. Briefly, each of the item nodes receives input for a short time after its corresponding stimulus is presented. This input activates the item node to a degree that decreases linearly with the number of item nodes already activated—a primacy gradient results. Alternatively, for those more familiar with models such as that of Burgess and Hitch (1992), the primacy gradient might be thought of as resulting from association of each list item with some representation of the start-of-list context, with the strength of association decreasing with list position (see later). Indeed, in the positional (i.e., negligible chaining) model that Burgess and Hitch favored, a primacy gradient of the type we require is generated simply by cuing with the pattern representing the first position (with a small change necessary to ensure all list items are activated above zero). The important difference is that, in this version of our proposed model, once the primacy gradient is established, ordered recall can proceed without any need to cue further positions, that is, without any need to "move the context along."

At first sight, the primacy gradient might seem to be at variance with the conventional assumption that recent memory representations will generally have a stronger trace than older representations that have been subject to decay or interference. That is, one usually expects information in memory to have a recency gradient and not a primacy gradient. The important thing to remember, however, is that the primacy gradient represents the activation of item representations specifically in the presence of a cue to serial recall. Even so, decay of activations over time is central to the proper operation of the primacy model.

The mechanism by which a primacy gradient is converted to an ordered response is simple. Once the primacy gradient is established, recall consists of a repeating cycle in which the item with the largest activation is selected for recall and subsequent recall of that item is suppressed. Suppression is assumed to be necessary to prevent repeated recall of the most active item. The model is neutral with respect to the exact mechanism of suppression, though this could be achieved in a number of ways (e.g., Burgess & Hitch, 1992; Houghton, 1990; Lewandowsky & Murdock, 1989). Clearly, if the primacy gradient is properly established and the maximally active node can be selected without error, performance will be perfect. To the extent that either the gradient itself or selection is contaminated by noise, recall will be prone to error.

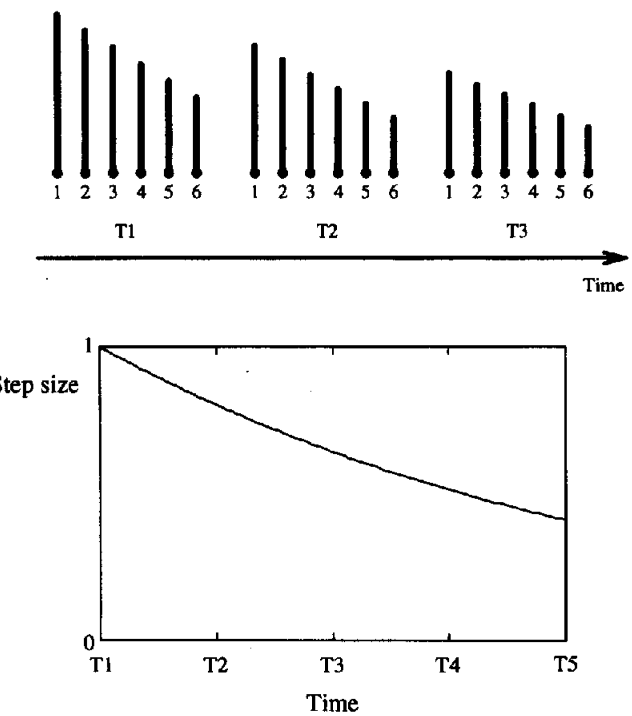

Figure 1. The effect of activation decay on the effective primacy-gradient step size. The upper section shows the exponential decay of primacy-gradient activations over time. The lower panel shows the corresponding exponential decrease in the effective step size, that is, the activation difference between nodes representing successive items.

Specific Mechanism

To make the mechanism of the primacy model entirely explicit, we assume that, in response to the stimulus list "123456," we have a primacy gradient of activations for which $x_1 > x_2 > x_3 > x_4 > x_5 > x_6 > 0$ and for which $(x_1 - x_2) = (x_2 - x_3) = (x_3 - x_4)$ and so forth. A constant-step primacy gradient is illustrated (at three different points in time—see later) in Figure 1. We further assume that although the primacy gradient is instated without noise, the process of choosing the item with maximal (unsuppressed) activation is noisy. (The fact that we concentrate the effects of noise into the choice process alone is purely for computational ease—the same results could be achieved by reducing the choice noise and assuming some additional noise in the primacy-gradient activations themselves.) In our model, noisy choice is modeled by adding zero-mean Gaussian noise to each of the activations before accurately choosing the node with the largest resulting activation. For simplicity, we assume that the primacy gradient itself remains unaffected by this noisy selection process, so that the addition of noise is not cumulative with regard to the original activations.

Data from ISR experiments indicate that the *primacy portion* of the serial-position curve, that is, the portion for which errors increase with list position, is more extensive than the recency portion, which is usually confined to the last position of recall. The primacy model explains the asymmetry by assuming that while recall is proceeding, the primacy-gradient activations undergo exponential, time-based decay. Thus, during recall, the effective difference between the activations of the yet-to-be recalled items in positions n and (n + 1) decreases and, accordingly, correctly ordered recall of these items becomes more vulnerable to the constant-variance noise. The effect of decay is illustrated in Figure 1. The upper panel shows how activations, representing the list "123456," decay over time. The lower panel shows how this also causes the effective step size to decrease over time.

To explain primacy effects, our model requires only the assumption that activations decay during recall. In fact, using the activation-based conception of our model, we assume that an item node's activation, which grows under the influence of input in the short period immediately after presentation of the corresponding item, decays exponentially at all times thereafter (including during recall). To avoid this leading to a recency gradient under conditions of long stimulus interonset interval (IOI), the degree to which a given item node activates on presentation of its corresponding item depends not only on the number of items already active but in addition on a modulating factor whose value itself decays exponentially with time measured since the beginning of list presentation. This aspect of the model ensures that a constant-step primacy gradient results even in cases in which the IOI is long. Differential equations for item-node activations, which guarantee a primacy gradient in this way, are given in the Appendix. If, alternatively, the primacy gradient is conceived as resulting from a start-of-list context cue, then decay can be thought of as reflecting an exponentially diminishing ability to reconstruct/maintain this cue as time passes.

On the basis of these decay assumptions, the strength of the primacy gradient at the start of recall (and more particularly the step size that separates the activations of successive items at that time) will depend on the time that has elapsed since current-list presentation began. We might, therefore, assume that the longer the stimulus IOI (i.e., the slower the list is presented), the worse will be performance. The fact that this result is not typically obtained has been attributed to participants' being able to use the longer stimulus IOIs to perform subvocal rehearsal, a proposal supported by data indicating that slowly presented lists are recalled more poorly under conditions in which rehearsal is prevented by articulatory suppression (Baddeley & Lewis, 1984). In our model, we assume that a cumulative rehearsal, in order, of items so far presented simply constitutes a more recent re-presentation of the list. In the presence of effective rehearsal, therefore, the strength of the primacy gradient at the start of recall will depend on the time that has elapsed since the start of the most recent rehearsal rather than on the time since the start of list presentation; this aspect of the model is illustrated in Figure 2. We suggest that rehearsal proceeds until the point in the list at which a full rehearsal of all the items so far presented becomes impossible in the stimulus IOI. Clearly, the longer the stimulus IOI, the later in the list cumulative rehearsals will be possible, though the longer will be the time delay, for a given list length, once rehearsal has ceased and before recall can begin. This aspect of our model is crucial in explaining effects of word length, list length, and delay.

The assumption that rehearsal in serial-recall tasks is *cumulative* in nature, as opposed to *repetitive* (in which only the most recently presented item is repetitively rehearsed) or *associative* (in which the two most recent items are rehearsed in order so as to cement their supposed association), receives some experimental support. Systematic investigation of participants' covert rehearsal strategies is difficult, but the results of experiments employing overt rehearsal or instructed covert rehearsal (Ferguson & Bray, 1976; Palmer & Ornstein, 1971) are consistent with the idea that cumulative rehearsal is both optimal and usual for serial-recall tasks.

The model as described so far is only capable of modeling order errors. It is important to note, however, that even this simple model, based on a simple primacy gradient of activations, is capable of showing the one-item recency effect found in the data. In fact, it also shows a one-item primacy effect, though this is often masked by a ceiling effect at early recall positions and by the more general primacy effect that results from activation decay during recall. These *end effects* come about as follows: The errors underlying the curve are overwhelmingly paired transpositions of adjacent list items, that is, items n and (n + 1) exchanging positions. Because of the noisy choice process, the activation of item (n + 1) will sometimes exceed the activation of item n, leading item (n + 1) to be recalled before item n. The first list item can only be involved in such an error with the second list item—it cannot be recalled too early. Likewise, the last list item can only be transposed with the penultimate item—it cannot be recalled too late. By contrast, items from the middle of the list can be exchanged by being recalled early in some trials or late in others. This leads to a situation in which, averaged over many trials, transposition errors on the first and last items are less probable than such errors on other list items. Analogous primacy and recency effects due to end effects are also seen in the perturbation model of Estes (1972) and that of Lee and Estes (1977, 1981).

Primacy and recency effects are natural consequences of the simple primacy model. However, early simulations revealed that this model consistently had a larger recency effect than that seen in the data. (The recency effect is measured as the difference between performance on the final and penultimate items, positive recency being reflected by better performance on the final item.) A detailed comparison of the real and simulated serial-position curves afforded a simple explanation. The recency effect predicted by the model actually showed a good correspondence with that found in the data if analysis was restricted solely to transposition errors. However, in the experimental data, the strong recency effect shown in transpositions is tempered by the occurrence of other types of errors, such as omissions, intrusions, and repeats. Collectively, these item errors tend to occur with a frequency that increases with output position, with no reduction for the last position (Henson et al., 1996). In the Henson et al. experiments, the number of omissions and intrusions was not large enough, even at the last position of recall, to mask the contribution of transposition errors to the overall last-item recency effect. In the light of this observation, we decided to extend the model to incorporate an account of such errors. We did this by introducing the possibility that, as activation levels decrease, item errors occur in addition to order errors.

The mechanism by which we model item errors and their increase with output position is as follows. After the item with the maximal activation (with noise) has been chosen, this activation is forwarded to an output decision process. Here, the forwarded activation is compared with a threshold (cf. Burgess & Hitch, 1992) to decide whether the corresponding item is either recalled (above threshold) or not (below threshold). In the latter case, either the item will be omitted or, in an experimental setting, guessing might result in an alternative item's being recalled in its place. Like the item-selection process, the threshold-comparison process is assumed to be subject to noise. Introduction of such a mechanism allows us both to simulate errors other than transpositions and to produce an improved fit to the observed serial-position curves.

The provision of an output threshold is intended to simulate the increase in item errors with output position. The primacy model, with its monotonically decreasing level of activation across list items, is able to account for this error pattern in a natural way, without sacrificing the one-item recency effect found in the pattern of transposition errors. Note that the suggested mechanism can also account for the observation that it is possible to recover from, for instance, an omission, by giving the correct item in the following position of recall. The pattern of errors found within the class of item errors, that is, the individual

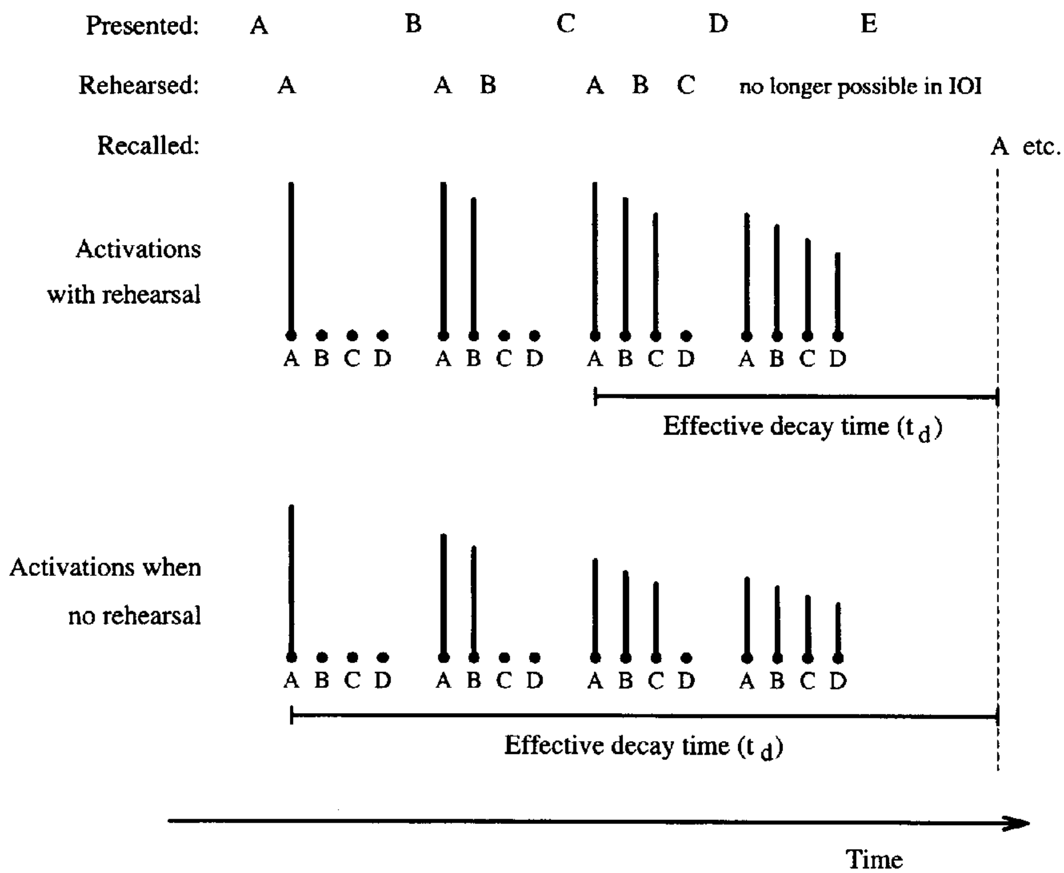

Figure 2. The effect of rehearsal on the level of primacy-gradient activations. Cumulative rehearsal acts like a re-presentation and delays the time at which decay starts having its effect. IOI = interoset interval.

patterns of omissions, intrusions, and repetitions, is not explicitly modeled here. Such explicit modeling would lead to the introduction of several additional parameters while offering, we believe, very little in the way of explanatory benefit over the qualitative account proposed next. We assume that omissions occur as described above but speculate that there are occasions when participants are unwilling to omit and, in preference, respond with an item. Occasionally, this item will not have been present in the stimulus list (though typically it will have been presented in a recent list), leading to an intrusion error. Repeat errors are potentially explicable in a similar way. Henson et al. (1996) characterized repeat errors as being literally few and far between, that is, the erroneous repetition of an item tends to be separated from its first occurrence in recall by several intervening responses. In Henson et al.'s data, repeat errors consisted chiefly of early items being recalled again in a late (predominantly the last) position of recall. In the present framework, this phenomenon is most readily explained in terms of "wearing off" of response suppression. As noted above, in our model, items that have already been recalled are necessarily suppressed to prevent their being recalled again and again. If this suppression were to weaken during the course of further recall, then one would expect items recalled early to begin to be recalled again at later positions. This is particularly the case if one envisages the primacy gradient itself to remain unaffected by recall in spite of the response suppression applied to recalled items. We do not model the wearing off of suppression here.

Simulations and Fits to Data

The primacy model has five parameters. Four of these parameters can be arbitrarily scaled and are thus defined relative to the step size in the primacy gradient as it would appear if the rate of decay was set to zero—this notional step size is therefore set to one for simplicity. N is the standard deviation of the selection noise, D refers to the exponent of the activation decay such that an activation of x at a given time becomes an activation of xe^{-D} 1 s later, P is the peak value of the undecayed primacy gradient, T is the output threshold, and M is the standard deviation of the noise added to the forwarded activation before comparison with this output threshold.

Before we could fit data from a given experiment, we needed to make some approximations regarding rates of covert rehearsal and output times for the experimental stimuli. Where there were published data available to help us make these approximations, we used them accordingly; where data were unavailable, we made what we believe to be plausible assumptions. The presumed values for various stimulus types are given in Table 1: The values for words were estimated from Baddeley et al. (1975), Baddeley and Andrade (1994), and Hulme et al. (1991); the values for letters were estimated so as to be consistent with these data. These values remained fixed throughout our simulations and were not manipulated to produce good fits.

Given these rates, the number of cumulative rehearsals, C, performed by participants was estimated as $C = R(I - 0.2)$, where I represents the stimulus IOI used in the experiment, measured in seconds, and R denotes the covert rehearsal rate. The constant 0.2 represents an approximation to the time needed to recognize the list item. Therefore, the time over which the primacy gradient is assumed to have decayed at the start of recall is given by $I \cdot \max(1, L - C)$, where L is the number of items in the stimulus list. Given approximations of the recognition time and rehearsal rates, this delay time can be calculated for any given experiment.

The first data we used for the fitting of the model to serial-position curves came from two sources: Baddeley (1968, Experiment 5) and Henson et al. (1996, Experiment 1).[3] In both cases, participants were asked to perform immediate written recall, in the order of presentation, of six visually presented letters. Henson et al. noted some minor differences in experimental design between the two experiments, chiefly concerned with the blocking or otherwise of lists of different types.

Our initial concern was to fit our model to the data from lists that contained only nonconfusable items. Initial optimization was performed both by hand (which is feasible for low-dimensional parameter spaces) and by using the nondeterministic minimization algorithm described by Caprile and Girosi (1990). As a result of this process, four of the five parameters were set as follows: D = 0.27, P = 11.5, T = 0.49, and M = 0.74. The remaining parameter, N, was, unless otherwise mentioned, the only parameter that was varied so as to fit all the data presented below. (In most cases, we have been able to improve on these fits by allowing more than one parameter to vary slightly between data sets, but the benefit does not, in our view, justify the consequent loss of clarity in presentation of the model.) All results shown were obtained by running the simulation for an equivalent of 10,000 stimulus lists, which was more than sufficient to ensure a stable pattern of results between runs.

The fits produced by the five-parameter model are shown in Figure 3. The noise parameter, N, was set to 0.19 for simulating the Baddeley (1968) data, giving a rms error of 0.032 and to 0.23 for simulating the Henson et al. (1996) data, giving a rms error of 0.017. The pattern of item errors produced by the model and that found in the Henson et al. data are also plotted in the left panel of Figure 3 (the equivalent data are not available for the earlier study). The model does a reasonable job of capturing the pattern of item errors, which are at a generally low level

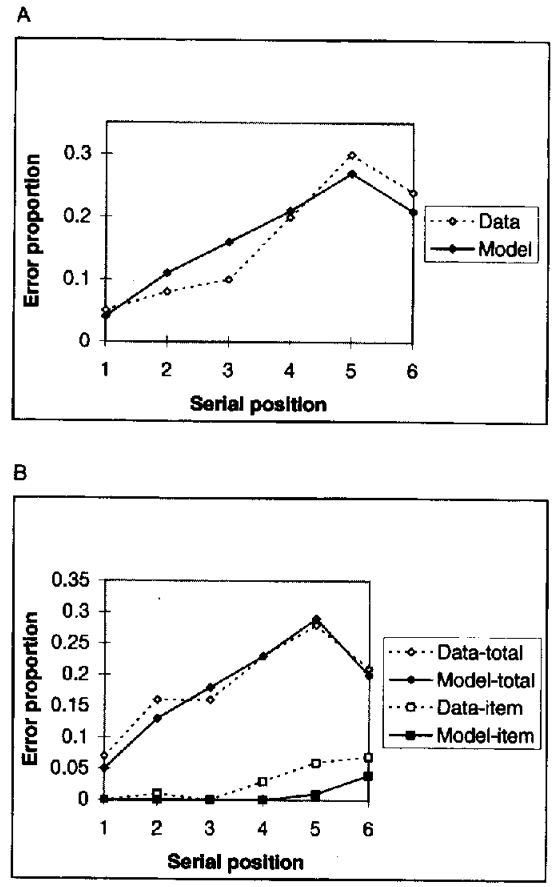

Figure 3. Simulations of serial-position curves for lists of nonconfusable items, using the five-parameter model. Panel A shows simulations of data from Baddeley (1968, Experiment 5). Panel B shows simulations of data from Henson et al. (1996, Experiment 1).

and which increase with list position, the rms error for the item-error curve being 0.027. In fact, early simulations of a two-parameter (N and D) primacy model, without an omission mechanism, showed that such a model was also able to fit the overall serial-position curve with an rms error of less than 0.04. The justification for the additional complexity of the extended model is primarily that it allows the modeling of a distinct type of error rather than conferring any drastic improvement in the ability to fit these data. In order to test whether the discrepancies between the data and the model are significant, we used Hotelling's T^2 test of the null hypothesis that the serial-position curves for model and data are statistically indistinguishable. We were able to do this for only the Henson et al. data because we required variance and covariance data that were unavailable for the earlier study. For the total and item errors, the F values, based on 6 and 42 degrees of freedom, were 0.37 and 0.08,

Table 1
Rates of Covert Rehearsal and Output Times Assumed for Simulations

Material	Covert rehearsal rate (items/s)	Output time (s/item)
Letters	4.0	0.5
One-syllable word	2.8	0.7
Three-syllable word	2.3	0.95
Five-syllable word	1.6	1.4

[3] Henson et al. briefly reported simulations of their data using an earlier version of the primacy model. This earlier version did not incorporate decay directly and therefore had a very restricted scope compared with the full model described here.

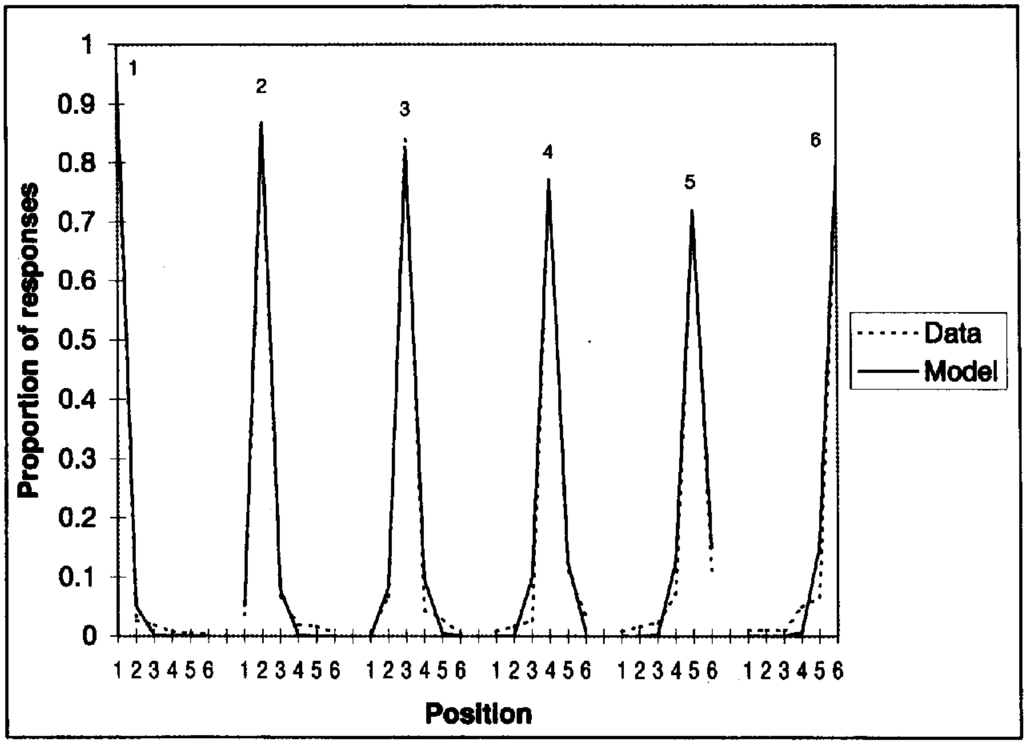

Figure 4. Simulations of transposition gradients from nonconfusable-list data from Henson et al. (1996), using the five-parameter model. The numbers above the curves refer to the ordinal position of the item in the stimulus list. The curves show the proportion of responses on which that item appears in each of the six response positions. For comparison with the model, the data do not include response repeats.

respectively, neither of which gives us any reason to reject the null hypothesis.

The transposition gradients for the Henson et al. (1996) data (again unavailable for the earlier study), showing the positions in which a given stimulus item is likely to be recalled, are shown in Figure 4. This figure shows that the model is consistent with the data in that the likelihood of a given transposition decreases with transposition distance (NB, this is not true of all models of immediate serial recall—e.g., Nairne, 1990). The rms error between the model and the data for the 36 points shown is 0.029.

Given that, as mentioned above, inclusion of the additional mechanism and parameters needed to account for item errors was motivated by the existence of this type of error in the data, rather than just a desire to improve the fits to the overall serial-position curves, it is desirable to show that the item-error mechanism also provides good fits to serial-position curves in which item errors form a larger proportion of the errors, relative to transposition errors, than they did in the Henson et al. (1996) study. Such data are difficult to find, however, because very few studies provide serial-position curves for item errors as distinct from errors in general. We therefore used data that we collected in one of our own experiments involving immediate serial recall of visually presented lists of eight letters, an experiment for which we have a full error analysis.[4] Figure 5 shows the model fit to the serial-recall curves for both total errors and item errors (note that the curve depicting order errors is totally constrained by these two curves and is therefore, for the sake of clarity, not shown). The value of the noise parameter, N, was 0.17. Once again, the fit is good: The rms error for the total-error curve is 0.028, and the rms error for the item-error curve is 0.029. In both cases, Hotelling's T^2 test, applied as above, gives $F < 1$.

It is worth pausing at this point to review the properties of

[4] These data were collected to investigate the pattern of errors underlying the modality effect. Participants performed immediate, written, serial recall of eight-item lists containing the letters *R, B, J, X, H, Z, L,* and *Q,* with no repeats. The lists were arranged so that no letter appeared in the same position as it had in either of the two previous trials and so that each letter appeared approximately equally often in each position. In the visual condition, the letters were presented one after another in the center of a computer screen, so that each letter replaced its predecessor. In the auditory condition, participants heard lists of the same letters through headphones. The speaking of each of the letters was digitally recorded, and the lists were assembled from these samples, with care taken that all lists sounded rhythmically isochronous. In both conditions, the presentation rate was one letter per second. Immediately after the presentation of the list, a recall cue appeared on the computer screen and participants wrote their responses, strictly from left to right, in a row of eight boxes. They responded in their own time and were asked to indicate any omissions by putting a stroke through the relevant box. They covered their response before proceeding to the next trial. In each of the conditions (visual and auditory), there were 18 participants, each of whom recalled 60 lists. Modality was treated as a between-subjects factor.

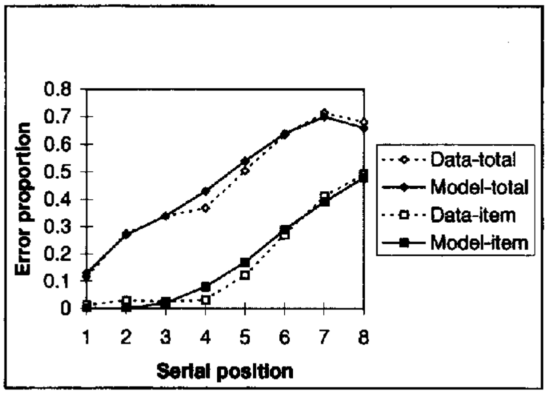

Figure 5. Simulations of serial-position curves for visually presented lists of eight nonconfusable letters, using the five-parameter model.

the primacy model that are responsible for its ability to give such an accurate simulation of serial recall. End effects generate symmetrical effects of primacy and recency restricted largely to the terminal items themselves. End effects are simply a consequence of the fact that the terminal items have fewer opportunities to engage in a transposition error than items in the middle of the list. Overlaid on top of these very local effects is the influence of decay, which has its impact throughout the entire list. The effect of decay is to reduce the effective step size for items recalled later in time (see Figure 1), making transposition errors more likely. At the very end of the list, end effects overcome the disadvantage of a decreasing step size to produce the upturn in performance we see as a recency advantage. Thus, somewhat paradoxically, the primacy model generates a recency effect even though the last item has the lowest initial level of activation and has the smallest effective step size between it and the preceding item. Recency is thus an automatic consequence of the general principles of the model and is not something that demands any special treatment of the final items in a list. This recency advantage in transpositions is opposed, but not normally overwhelmed, by a tendency for item errors to increase with output position. These item errors are modeled by the inclusion of an output threshold for responses.

A closer look at order errors reveals that the primacy model has a tendency toward local transpositions because of a property that we term *fill-in* (Norris, Page, & Baddeley, 1994). The property itself can be simply stated: When an item is missed in recall, due to a local transposition, it is liable to be recalled in the next position. For example, in response to the list "123456," if Item 2 is recalled in the first position, then Item 1 is most likely to be recalled next. Analysis of the Henson et al. (1996) data shows that this is indeed a property of actual recall, Item 1 being recalled three times more often than Item 3 in cases in which Item 2 was erroneously recalled first. Fill-in is obviously a property of the primacy model, because the activation of early items always exceeds that of later items, provided neither has been recalled and thus suppressed. It should be noted, however, that fill-in is not naturally a feature shared by either item–item chaining models or position–item association models. As a concrete illustration of the latter, we shall refer to the positional model suggested by Burgess and Hitch (1992) toward the end of their article. The aspect of this model that we wish to highlight is the fact that the positional codes assumed to underlie recall cue items symmetrically about the correct position. That is to say, a given context vector maximally cues the recall of the item with which it was associated at list presentation, and it cues adjacent items equally. Translated into the activation framework presented here, positional cuing of the second item in a six-item list would lead to the symmetrical pattern of activation seen in Figure 6, in which items from the first and third positions are cued equally. (Note that in position–item association models, the identity of the item or items previously recalled has no effect on the cue for later items.) This pattern of response cuing has the following consequence: If, as before, Item 2 is recalled in the first position, then the next context state will cue Items 1 and 3 to an equal extent. Thus, there will be a 50% probability of recalling Item 3 in the second position (if we assume for simplicity that we can neglect all but one-apart transpositions). If Item 3 is thus recalled, the probability of recalling Item 1 decreases rapidly: The third context state is a much stronger cue for Item 4 than it is for Item 1 (Items 2 and 3 being suppressed). If response suppression is absolute, it is likely that Item 1 will not be recalled until the last position of recall, at which point it is the only response left unsuppressed. This not only reduces the chances of achieving a recency effect, which requires that the last item be preferentially recalled in the last position, but also results in an unusually large number of long-distance transpositions, in this case Item 1 transposing to Recall Position 6. Thus, it appears that the experimental data are incompatible with any positional model that cues items from adjacent positions symmetrically or in a manner biased towards succeeding items. These data are also incompatible with models relying on pairwise item–item chaining. Such models predict that if Item 2 is recalled in the first position, Item 3 will tend to be recalled next. This is the exact opposite of fill-in.

The Word-Length and List-Length Effects

Having shown that the primacy model performs well in simulating the basic properties of serial recall, we now demonstrate

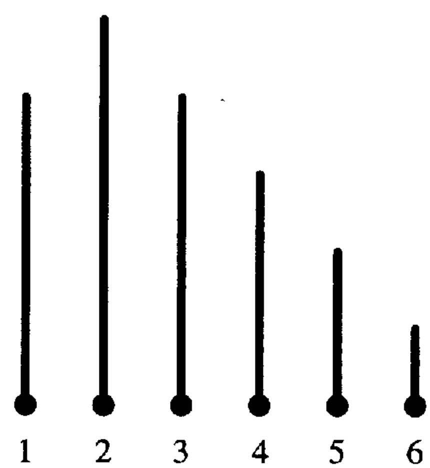

Figure 6. The activations of items when cued with a solely positional context corresponding to the second position of recall of a six-item list. This pattern assumes symmetrical cuing.

how this simple model can account for other characteristics of immediate serial recall—in particular, the word-length and list-length effects.

As noted above, the most interesting property of the relationship between word length and performance in serial-recall tasks is the fact that span appears to be linearly related to the rate at which the list items can be articulated (Baddeley et al., 1975). Various hypotheses have been proposed to account for this finding. Baddeley (1986) suggested that items in the phonological store are subject to decay unless periodically refreshed by a (typically subvocal) articulatory rehearsal process. Because long words take longer to rehearse than do short words, lists of long words permit fewer opportunities for rehearsal and thus their memory representations are subject to increased decay. Cowan and colleagues (Cowan, 1992; Cowan et al., 1992; Cowan, Wood, & Borne, 1994), although accepting the generality of Baddeley's account, further suggested a role for output delay. They postulated that memory decay continues throughout recall, and that, as a result, recall of items reported late in the serial recall of long-word lists will suffer relative to the equivalent items from short-word lists.

Note that if the length effect were entirely attributable to decay during output, there should be no effect of length on the first item in a list. Such an effect is typically found (e.g., Avons, Wright, & Pammer, 1994; Baddeley et al., 1975; Baddeley, Lewis, & Vallar, 1984) but was absent from Cowan et al. (1992). (One likely reason for this absence is that Cowan et al.'s results were restricted by a ceiling effect for the first item; another is discussed later.) The primacy model assumes a decay process that occurs during both presentation and output. Our assumption that decay has its effect at all times after cumulative rehearsal has ceased, and that that moment comes sooner for lists of longer words due to restrictions on the time available for cumulative rehearsal, is consistent with both Baddeley's (1986, etc.) and Cowan's (1992, etc.) accounts of the word-length effect. It is also consistent with Dosher and Ma's (1998) finding that output time is a crucial determinant of ISR performance.

To investigate the behavior of the primacy model with respect to word length, we tried to simulate the linear relationship between span and articulation rate. There is considerable variation in the literature regarding the ways in which both span and articulation rate are measured. For instance, Baddeley et al. (1975) measured span using the "mean percentage of words recalled in the appropriate position" (p. 582) for a list of fixed length. The more usual definition, and the one we simulate here, gives span as the list length for which 50% of lists are recalled correctly. The comparison data came from the first experiment in Hulme et al. (1991). The simulations were run with the noise parameter, N, set to 0.16 and the rehearsal and output rates given in Table 1.

The results of the word-length simulations are shown in Figures 7 and 8. Figure 7 shows the variation in proportion of lists recalled correctly with list length, for three different word lengths (viz., one, three, and five syllables). The characteristic inverse S curves demonstrate that the proposed mechanism captures the list-length effect and lead to estimates for the 50%-correct span. These span estimates are plotted in Figure 8, together with corrected[5] spans from Hulme et al. (1991). The rates for speeded articulation for the experimental data were estimated with reference to Baddeley et al. (1975). The simulations not only give the desired linear relationship (R^2 = .99, slope = 1.2, intercept = 2.7) between span and articulation rate but also match the experimental results to well within the experimental error that can be inferred from the Hulme et al. article (slope = 1.4 with a standard error of 0.19; intercept = 2.4 with a standard error of 0.27).

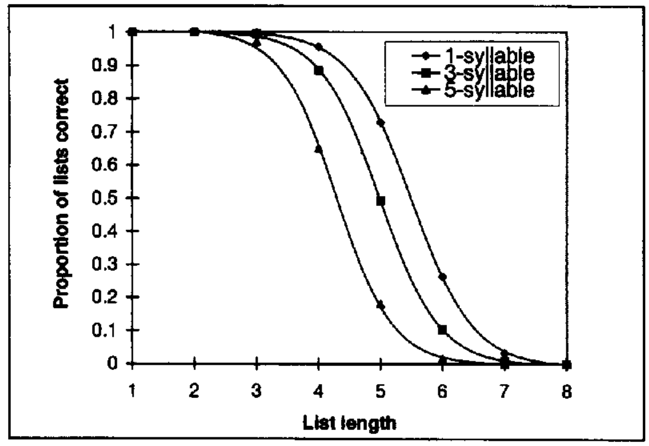

Figure 7. Simulations showing the variation of percentage of lists correct versus list length for a variety of word lengths. The 50%-correct span measure is derived from these curves.

There is one further interesting point to make about these word-length experiments. As was noted earlier, Cowan et al. (1992) found an interaction between word length and serial position: There was no effect at the first position of recall, but the effect increased with list position. It was hypothesized above that this was partly due to ceiling effects for early items. The model can shed further light on this finding, because it makes explicit the contributions of prerecall decay and during-recall decay. The Cowan et al. experiment used a comparatively long stimulus IOI of 2 s. On the basis of the assumption used earlier, this IOI allows approximately 1.8 s available for cumulative rehearsal in the time between successive stimulus items. Both the short and long words consisted of two syllables, differing only in articulation time. Rates of speeded articulation, inferred from Cowan et al., suggest that for both long and short words (2.27 and 2.63 words per second, respectively), more than four words could be articulated in an interstimulus interval. The lists themselves were five items long, so "full" cumulative rehearsal

[5] The correction involves adding 0.5 to the span measures obtained experimentally. This is because Hulme et al. (1991) calculated span by giving participants four lists at each progressive list length; span was determined as the list length at which participants performed four lists perfectly, plus 0.25 for each longer list correctly recalled. Thus a participant who got all lists of length four correct, two lists of length five, and none of length six would be accorded a span of 4.5. The correction between span measures is needed because, on the 50%-correct span measure, the same participant would be accorded a span of 5.0. This correction is valid on average as long as one assumes that the inverse S curve relating percentage of lists recalled versus list length is rotationally symmetric about the 50%-correct point.

is possible for both word lengths. Under these conditions, there is no difference between the prerecall decay in the short-word and long-word conditions. Thus, in this experiment—and, we predict, in other experiments combining short lists with long interitem intervals and relatively short word lengths—any word-length effect is bound, according to our model, to reflect only memory decay during output. In these circumstances, there will be no effect on the first item of recall, with the effect increasing with list position. To illustrate the point, we set up an approximate simulation of the Cowan et al. experiments. For the all-short and all-long conditions, we averaged the results across the first two experiments and adjusted the model's noise parameter, N, so as to optimize the model's fit with respect to values of the mean proportion of items correct given by Cowan et al. (short, .865; long, .785).[6] With N = 0.23, the model gave values of mean proportion correct of .86 and .79 for lists of short and long words respectively, giving the serial-position curves shown in Figure 9. The serial-position curves show exactly the pattern found in the data, namely, a ceiling effect at the first position, which contributes to an interaction between the effects of word length and serial position. It should be noted that these simulations necessitate assumptions concerning the time taken for participants to output the short and long words. The time taken to output a short word was set to 0.8 s, a value derived from other studies in which similar durations were measured directly. The time assumed for output of a longer word was then varied so as to achieve the best fit to Cowan et al.'s summary data. The best fit was achieved when this time was set to 1.1 s, which seems to be reasonable, but which, in the absence of experimental data, must remain without confirmation.

We have shown that a simple decay process, partially offset by cumulative rehearsal, is sufficient to account for the word-length and list-length effects in immediate serial recall of lists of familiar items. We have also illustrated how the detailed pattern of results can depend on the experimental design. G. D. A. Brown and Hulme (1995) also suggested that the word-length effect can be explained using trace decay. Theirs is not a process model but rather produces recall probabilities directly. It is difficult for such models to explain the types of error underlying the changes in span that they simulate. Indeed, aspects of their model imply that changes in span result solely from changes in the rates of omission errors. In contrast, the primacy model is consistent with the data, in that the adverse effects of increased word-length and increased list-length result from increases in both transpositions and item errors. The Burgess and Hitch (1992) model uses decay of position–item associations to achieve a word-length effect without investigating any contribution of rehearsal. Rehearsal is clearly an important part of the phenomenology of immediate serial-recall tasks, although it is necessarily rather difficult to control experimentally and to model. We believe that the primacy model represents an appropriate balance between rehearsal and decay processes that can simulate, both qualitatively and quantitatively, a variety of data.

The primacy model also affords a qualitative account of the interactions that are found between the word-length effect and articulatory suppression. The fact that the word-length effect disappears when presentation is visual and articulatory suppression is employed, even if only during presentation (Baddeley et al., 1975), can be explained by assuming (with Baddeley, 1986) that suppression prevents the visually presented material from being recoded so as to gain access to the phonological loop. With auditory presentation, the word-length effect persists if suppression is required only during list presentation and recall is vocal (Baddeley et al., 1975). In this case, suppression prevents rehearsal but does not prevent differential output delays from generating a word-length effect. Finally, when presentation is auditory, and suppression is in addition required throughout recall, the word-length effect is abolished (Baddeley et al., 1984). We can account for this result by noting that Baddeley et al. (1984) not only prevented rehearsal, with the requirement

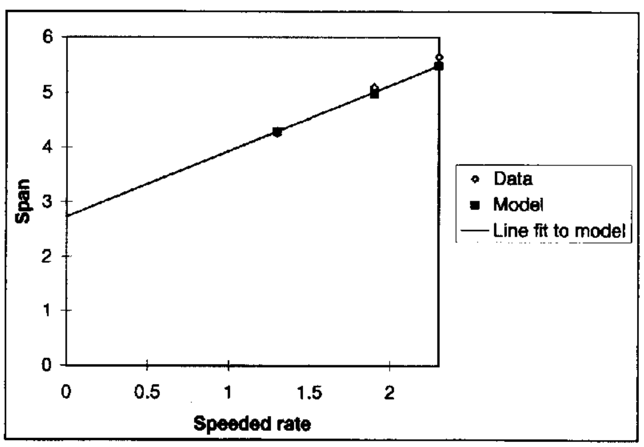

Figure 8. Results of simulations showing the linear relationship between 50%-correct span and articulation rate, compared with corrected data from Hulme et al. (1991). Both panels refer to the same simulations as were used to generate Figure 7.

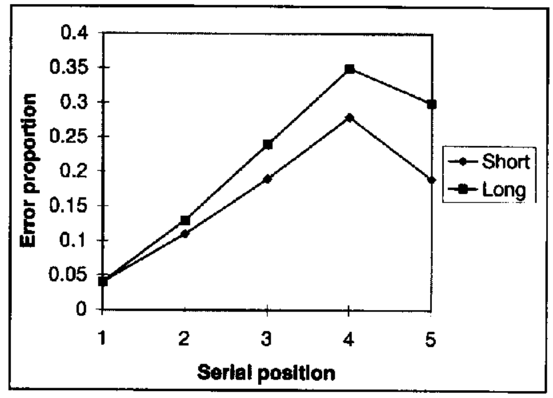

Figure 9. Simulation showing how the word-length effect can interact with serial position. The upper curve simulates performance with long words, and the lower curve simulates performance with short words, under those conditions inferred from Cowan et al. (1992).

[6] Exact fits to the serial-position curve are not possible in this case, because Cowan et al. did not give the relevant information for both experiments.

to suppress throughout, but also deliberately attempted to equate written output times for the long and short words by allowing participants to abbreviate the former. Under these conditions, the primacy model is consistent with the data in predicting no effect of word length.

Neath and Nairne (1995), by building on Nairne's (1990) feature model, suggested that word-length effects can be observed in a model that does not use the concept of decay. Their alternative is to postulate a "theoretically neutral" (p. 432) property of a word, namely, its "number of segments" (p. 432), that is perfectly correlated with its articulatory duration. In addition, they assumed that words with more segments suffer at recall in a probabilistic segment-assembly process. The interesting feature of Neath and Nairne's model is that the only motivation for these assumptions appears to be to avoid attributing the length effect to decay. Rather than accept that increased word length leads to extra decay, they introduced two extra features (i.e., segments and an unspecified segment-assembly process), neither of which explains anything that is not already explained by decay-based models. Moreover, as they noted, it seems difficult to square their account with the data of Cowan et al. (1992), who found that recall of short words was adversely affected if their recall was preceded by recall of long words from the same list.

We can also apply the primacy model to data from delayed-recall tasks in which rehearsal is prevented during the retention interval. J. Brown (1958), Conrad (1958, 1960), Murdock (1961), and Peterson and Peterson (1959) all showed that immediate recall could suffer greatly even from a short filled delay. The effect was particularly apparent for lists of length approximately equal to span. Conrad (1960) found that simply requiring participants to prefix their responses with a redundant item took their performance from 73% of lists correct to 38% of lists correct. In terms of the primacy model, we can relate this to the steep portion of the inverse S curves showing recall performance plotted against list length. When participants are operating at around span, a single extra item, or in this case a single extra unrehearsed delay, causes a drastic drop in performance. Dallett (1964) supported this interpretation by showing that a seven-digit list prefixed with a redundant "0" either on presentation, on recall, or both, was recalled at the same level as an eight-digit sequence with no redundant elements. (Note that Dallett interpreted the result of the condition involving prefixed presentation with nonprefixed recall as indicating that participants found it impossible not to remember the prefix as part of the list and therefore took time to omit it from their recall.) A more detailed investigation by Dallett (1965) and another by Crowder (1967) both reinforced the conclusion that a redundant item in recall has an effect very similar to lengthening the to-be-remembered list by one nonredundant item. Baddeley and Hull (1979) similarly attributed the prefix effect, and the greater part of the suffix effect, to decay over an interval during which rehearsal is prevented, and this view was also supported by data from Mortenson and Loess (1964), Lowe and Merikle (1970), and Jahnke (1975).

A final observation, related to the issue of recall delay, concerns backward recall. If participants are required to recall the stimulus list in the reverse direction, the typical finding (e.g., Hulme et al., 1997; Madigan, 1971) is that errors increase with output position, with a small benefit for the last items recalled (i.e., the first items presented). We speculate that, when faced with this task, participants are able to respond quickly with the last stimulus item but thereafter proceed with a number of covert forward recalls of decreasing length, each time "peeling off" the last item for recall. Some evidence (Anders & Lillyquist, 1971) suggests that participants are able to peel off pairs of items and to use visual or other strategies to perform their reversal. (This suggestion certainly accords with our subjective experience when performing the task.) We used the model to simulate this account, assuming that participants peeled off items in pairs. The noise parameter, N, was set to 0.18, and the data were those taken from Hulme et al. (1997, Experiment 4). In fact, Hulme et al. also varied the frequency of the words used to test forward and backward serial recall. Consistent with their proposal that word frequency has its effect at a late *redintegration* stage, we varied the omission threshold, T, in our model accordingly: T was left equal to 0.49 for high-frequency items and was set to 0.90 for low-frequency items. The only further assumption we made concerned the time taken to reverse the last two items following a covert forward recall—we estimated this to be 2 s, although no timing measurements were made that could confirm the accuracy of this estimate.

The serial-position curves for forward and backward recall, showing errors plotted against input position for both the data and the model, are shown in Figure 10. Given the assumptions that we made regarding the strategy used by participants in this experiment and the timing of their responses, it would be inappropriate to make quantitative claims on the basis of our simulations. Nonetheless, a number of points can be made regarding the model's qualitative performance. First, the form of the serial-position curves produced by the model is close to that seen in the data. (Note that the divergence between model and data in the last position of forward recall is attributable to a modality effect, the experimental lists having been presented auditorily—see later.) Moreover, the model shows the correct pattern of results with respect to word frequency: There is an effect of frequency in forward recall that grows over serial positions, but there is no such effect apparent for backward recall. This pattern of results emerges because the model shows fewer omissions for backward recall than for forward. The repeated covert forward recalls act like rehearsals, which, although they might introduce cumulative order errors, maintain primacy-gradient activations at a reasonably high level relative to the omission threshold. Backward recall is thus less sensitive to a change in this threshold than is forward recall. The performance of the model with respect to these data simultaneously supports both our general approach to backward recall and our decision to model frequency effects by varying, on an item-by-item basis, the level of the omission threshold.

The Phonological Similarity Effect

As noted earlier, lists of items that are phonologically similar are typically recalled worse than lists of phonologically dissimilar items. Superficially, it may seem difficult for the primacy model, with its localist representation of list items, to capture this effect. Nonetheless, we show that a simple extension to the model enables simulation of data that has proved extremely

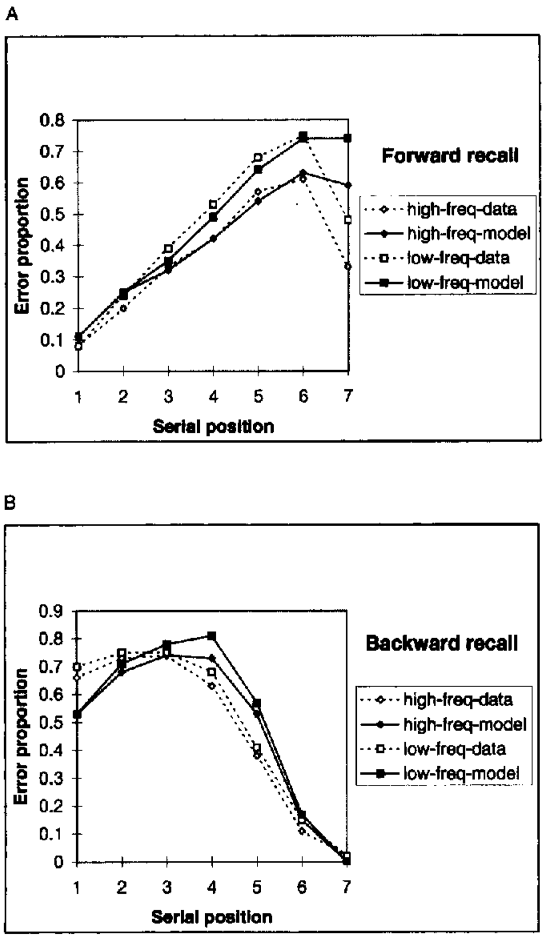

Figure 10. Simulation showing the effect of word frequency on forward and backward recall. The data are taken from Hulme et al. (1997). High-freq = high-frequency; low-freq = low-frequency.

difficult for alternative models and that has become something of a benchmark test for such models. These data concern recall performance for mixed lists, containing items of both low and high confusability. Such lists were used by Baddeley (1968) and by Henson et al. (1996), the latter extending the original results and augmenting them with a detailed analysis of the pattern of recall errors. Both Baddeley and Henson et al. found that recall of nonconfusable items in mixed lists was unaffected by the presence of confusable items in spite of the fact that performance on the confusable items suffers considerably. This leads to characteristic sawtooth-shaped serial-position curves for lists in which the items alternate in confusability, with the troughs corresponding to the nonconfusable items lying on the same curve as items from lists entirely composed of nonconfusable items.[7] Careful inspection of the errors underlying these serial-position curves reveals that the additional errors found for confusable items in mixed lists consist almost entirely of transpositions between the phonologically similar items (Henson et al., 1996). Concern that participants in these experiments might have used some predictability in the structure of the lists to aid recall is not supported. Baddeley presented his six different list types in a single block in random order; Henson et al. blocked the two types of alternating lists together but still presented them in a random order. In the latter experiment, participants were debriefed and claimed to have noticed no structure within or between trials. Moreover, Bjork and Healy (1974) also used mixed lists, in their case consisting of four items, two confusable, and presented all 24 possible orderings for serial recall in three delay conditions. Consistent with the findings of Baddeley and Henson et al., they concluded that

> the presence of two acoustically similar letters in the same to-be-remembered stimulus does not increase the loss of order information for the other acoustically dissimilar letters in the string, it seems to produce rapid loss of order information specific to the two similar letters. (p. 95)

These results suggested a two-stage mechanism (Norris et al., 1994; Norris, Page, & Baddeley, 1995), whose properties are characterized in Figure 11. The first stage can be considered to store the order of items in a manner unaffected by their phonological similarity. This function can be carried out by the model as described so far. The second stage can be considered to be a confusion stage. We loosely characterize this stage as an "output stage," though we are not committed to a particular interpretation of this label. The two stages interact as follows. At a point in recall, an item is selected at the first stage using the mechanisms described earlier. This single item is forwarded to the second stage, in which items' phonological forms are explicitly represented, where there is a chance that it will be confused with a similar sounding item (usually from the list—see later) before being output. The second stage is an additional source of transposition errors between confusable items, over and above the order errors, which are inherent in the first stage. A two-stage model with these general properties will necessarily produce sawtooth-shaped serial-position curves for alternating lists of confusable and nonconfusable items.

More specifically, recall is characterized in the following way. A single item at a time is chosen from the first stage as before. Localist item representations in the second stage are then activated to a degree dependent on their similarity to the item chosen at the first stage. The node corresponding to the forwarded item has activation 1; the nodes corresponding to items phonologically similar to the forwarded item have activation S, where S is a new parameter such that $0 < S < 1$; and nodes corresponding to items phonologically dissimilar to the forwarded item have zero activation. Next the activations are multiplied by the corresponding primacy-gradient activations as found in the first stage. The reason for this multiplication is detailed below. Then the item with maximal resultant activation is chosen by a noisy selection process entirely analogous to that in the first stage. The winning item is output, and its subsequent choice at the second stage is suppressed. In order to incorporate this mechanism into the existing model, we simply assume that this second stage is that stage at which the noisy threshold-comparison process (i.e., that which permits modeling of item errors) occurs. The only change to the model is the introduction of the single parameter, S, that allows phonological confusions, as well as item errors, to occur at this output stage.

[7] To aid in visualizing this effect, refer forward to Figure 12.

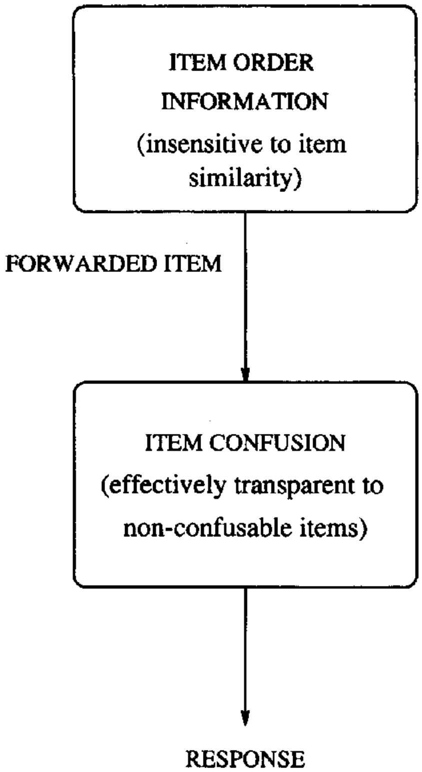

Figure 11. A schematic account of the proposed two-stage model.

Before showing the results of simulations, we must expand on two points in the preceding description. The first concerns the multiplication of second-stage activations by the corresponding first-stage (primacy-gradient) activations. At first glance, this might seem to be an unnecessary complication. Why should both stages be influenced by the primacy gradient? To appreciate the need for the primacy gradient to operate at both stages, consider the behavior of a simpler model in which the gradient operates only at the first stage. Once an item is chosen at the first stage, confusions at the second stage will be determined solely by the similarity between the item chosen and the other items remaining to be recalled. If a confusion error is made at the second stage, it will be equally likely to result in output of any one of the incorrect confusable stimulus items. In other words, the additional transpositions due to the confusion stage will be evenly distributed. This is not the case. For instance, Henson et al. (1996) showed that transposition profiles for the first item of recall for the pure confusable and pure nonconfusable lists are not parallel (i.e., related by an additive constant) as would be expected if the confusion stage were flat. The data suggest that the likelihood of a given confusion depends on the list positions of the items involved. We have chosen to model this pattern with the multiplicative mechanism described above. This also ensures that phonological confusions will result in transpositions rather than intrusions as is also found in the data.

The second point for expansion concerns the suppression of previous responses. The model is such that response suppression acts independently at each stage. Thus a confusion at the second stage will result in one item's (i.e., the item forwarded from the first stage) being suppressed at the first stage, with another item's (i.e., the item recalled) being suppressed at the second stage. To see why this should be so, imagine the situation that would result if the recalled item were suppressed at both stages, regardless of the item originally forwarded to the second stage. For example, assume that the alternating list "BRCX" has been presented, and that at the first position of recall the item "B" is chosen at the first stage and accordingly forwarded to the confusion stage. Further assume that a confusion occurs, so that the item finally recalled is "C," resulting in the suppression of "C" at both stages. In these circumstances, it is likely that "B," which is unsuppressed at either stage, will once again be chosen at the first stage in the second position of recall. This would result in a large number of additional errors in a nonconfusable position, an outcome that is at odds with the data. Of course, the fact that "B" is most likely to be chosen again in the second position of recall is a consequence of the primacy gradient continuing to favor the choice of "B" over "R." Would independent suppression at each stage prove unnecessary if, for instance, a positional cuing model were used, which would favor the recall of "R" over "B" in these circumstances? We propose that one can only dispense with independent suppression if one has a model that does not exhibit fill-in. Fill-in, as noted earlier, is the tendency to respond preferentially with previously skipped items; to the extent that this tendency is present, "B" will be recalled rather than "R" in the second position of recall in the above example, leading to increased errors in the nonconfusable position. Because fill-in is an essential property, we are forced to postulate independent suppression.

We used the model described above to simulate the serial-position curves from Baddeley (1968, Experiment 5). We employed the same values for the five original parameters as were used to simulate the serial-position curves for pure-nonconfusable lists, and we adjusted the additional parameter, namely, the degree of similarity of similar letters, S, to produce the other curves. This value remained fixed at S = 0.86 for all simulations presented here. Figure 12 shows the simulations, using these parameters, for the three list types CCCCCC, NCNCNC, and

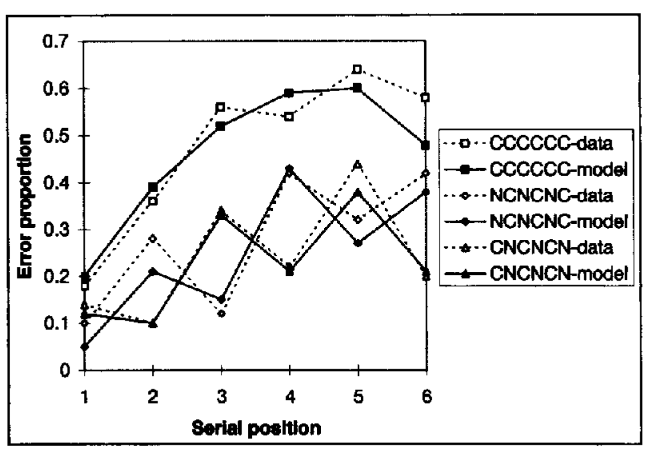

Figure 12. Simulations of three conditions from Baddeley (1968, Experiment 5). The key indicates stimulus-list composition, where N refers to a nonconfusable item and C refers to a confusable item.

CNCNCN (where N represents a nonconfusable letter and C a confusable) together with the corresponding data from Baddeley's experiment; the values for the rms error are 0.053, 0.045, and 0.026, respectively. Figure 13 shows the simulations for list types NNNNNN, CCCNNN, and NNNCCC, giving rms error values 0.033, 0.050, and 0.074, respectively. The comparatively high value for the rms error in the NNNCCC condition is largely attributable to a poor fit to the fourth and sixth positions of recall. There is a suggestion that participants were able to use a grouping strategy in this condition, which might help to explain these disparities. The model captures all the pertinent features of the data, particularly the insensitivity of the performance on nonconfusable letters to the confusability or otherwise of the other letters in the list.

The choice of a two-stage process to account for the effects of phonological similarity might strike some as inelegant. However, it is a choice that has been dictated by the complexities of the data. On the assumption that a one-stage model would be more parsimonious, we devoted a great deal of time to an attempt to develop a simpler account of the data. However, as in Burgess and Hitch (1992), none of the one-stage models we tested were able to give a proper simulation of the data. The preceding analysis, demonstrating the necessity for two stages of suppression to model the alternating-list data, highlights the central problem faced by one-stage models. We believe these data force the use of a two-stage model.

Is the notion of a two-stage process plausible? In many ways, the second stage can be thought of as similar to the deblurring process postulated to occur at the output of the theory of distributed associative memory model (Lewandowsky & Li, 1994); a two-stage process to model phonological similarity effects was also hinted at, without being fully specified, in Lee and Estes (1977). Furthermore, there are distinct similarities between the structure of our two-stage model and recent models of speech production. For instance, Levelt (1989) discussed in detail a two-stage view of lexical access for output. The first stage involves selection of a single lemma for output. The second stage involves activation of the phonological form. In this second stage, there may be activation of phonologically related items but not semantically related items. This is a precise parallel with our model: We assume the selection of a response at a level at which item representations are unaffected by phonological similarity, followed by access to the phonological representation of that item, with the possible output of phonologically related items (i.e., similar sounding letters) but not of semantically related items (i.e., other letters in general). In Page and Norris (1998), we made the link with models of speech production (in this case, those of Dell, 1986, 1988) more explicit, developing the idea that the phonological similarity effect results from the same process as do speech errors in everyday speech (cf. A. W. Ellis, 1980).

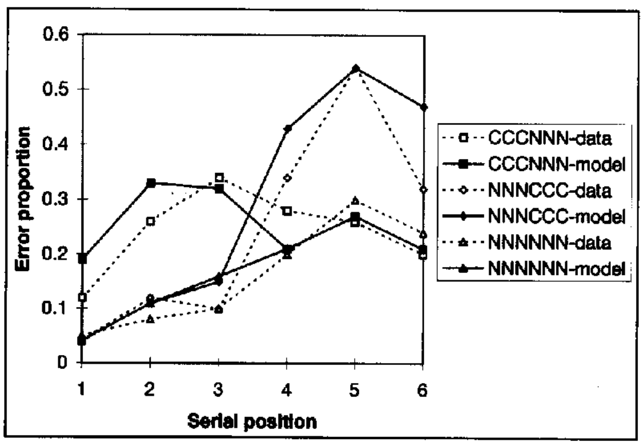

Figure 13. Simulations of the remaining three conditions from Baddeley (1968, Experiment 5). The key indicates stimulus-list composition, where N refers to a nonconfusable item and C refers to a confusable item. The NNNNNN condition is the same as that shown in Figure 3 and is included for completeness.

The two-stage idea also receives some support from data presented by Drewnowski and Murdock (1980), who identified two types of intrusion error in serial recall: one in which the intruded item is an item from a recent list but with no phonological similarity to the correct item which it replaces, the other in which the intruded item does not come from a recent list but is phonologically similar to the correct item. These intrusions, similar to those noted in normal speech by Garrett (1980), have a clear interpretation. The former result from selection of the wrong item at the first stage (though we have not modeled intrusions here); the latter result from selection of the correct item at the first stage, followed by a phonological confusion at the second stage.

A final observation cautions against identifying the phonological confusion stage with an articulatory output stage, namely, the finding that individuals with anarthria and those with dysarthria exhibit a normal phonological similarity effect (Bishop & Robson, 1989).

The Modality Effect

The final effect we discuss is the modality effect, whereby lists presented in the auditory modality are recalled better than those presented in the visual modality, the advantage being manifested in better recall of the list-final item or items. Our primary goal was to concentrate on modeling those effects on immediate serial recall, described above, that are common to both presentation modalities (cf. Crowder, 1978, p. 505). Indeed, our reading of the literature leads us to believe that the modality effect involves a component of memory distinct from the operation of the phonological loop itself. With this in mind, we present a simple simulation that illustrates that the modality effect need not rely on any better memory for order when presentation modality is auditory. Instead, we suggest that the effect relies on a better memory for items, or more particularly the list-final item, under these circumstances.

Suppose that auditory presentation of a list of items leads to a situation in which the last item heard is retained by the listener, in a way in which the last item seen in a visually presented list is not. We are completely agnostic about the mechanism underlying this memory, but note that such a memory for the last item heard involves no memory for order, merely memory for a single item. This account is, of course, very similar to that involving the precategorical acoustic store suggested by

Crowder and Morton (1969) and later modified by, for example, Morton (1976) and Crowder (1978). What consequences would such an additional memory have for serial recall of the list? Obviously, it will improve the recall of the last item, by anchoring it to the last response position: The listener knows the identity of the last item heard and "saves" it for the last position of recall. As a result, performance on the penultimate item will also improve, because there will no longer be a temptation to recall that item in the last position of recall (cf. the end effects mentioned earlier). In this way, a proportion of the improvement on the last position of recall can trickle back to the penultimate and, perhaps, the antepenultimate position.

To illustrate this point, we took some data that we had collected specifically in order to investigate the modality effect. The experimental method was absolutely standard and was described with reference to the simulations of eight-item lists presented earlier. We ran the model as before, with exactly the same parameters as were used in the earlier simulation, but for the auditory case we incorporated a new parameter, namely, the probability that the identity of the last item presented auditorily was retained uniquely in a separate memory store and was therefore bound to be recalled in the last recall position—and in that position only. This probability was set to .8 to give a good fit to the experimental data: The rms error with respect to the data from the auditory condition was 0.032. The experimental data and the simulation results are shown in Figure 14. Clearly, the provision of this single, largely reliable item memory, when accompanied by the basic ordered-recall mechanism, is sufficient to give an accurate account of the modality effect. Of course, we do not wish to place too much emphasis on the quality of the fit obtained here: The ability to tune the extra parameter allows us to fix the level of performance on the last recall position. Our purpose in including this section is merely to emphasize that the modality effect does not necessarily depend on any additional memory for order.

Is it likely that the last item heard should be retained in a way in which the last item seen is not? The question is not, perhaps, as well-specified as it might seem. When recall is

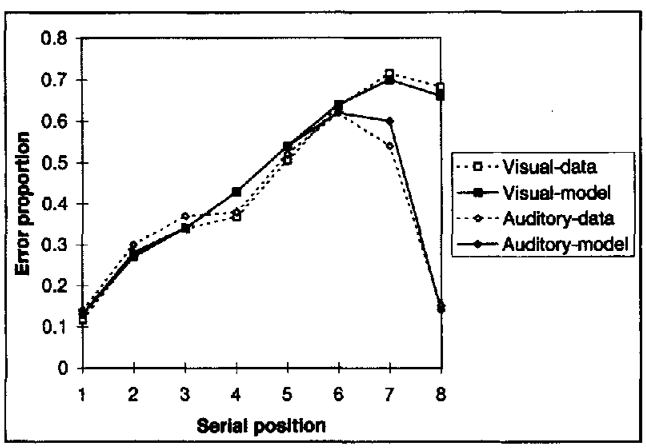

Figure 14. Simulations of the modality effect. The parameters are the same as those used for Figure 5, with an additional item-memory component assumed for the auditory case as described in the text.

written, as it is in most experiments designed to illustrate the modality effect, the last stimulus item presented at the time its recall is required really is the last item heard. By contrast, the last item seen is likely to be the response most recently written, rather than the final list item presented. Indeed, if participants are required to recall verbally, then the modality effect is drastically reduced and even disappears, consistent with the observation that, in this case, the last item heard will be the participant's own enunciation of his or her most recent response. A similar situation is evident when an irrelevant auditory suffix is appended to the stimulus list: The suffix now becomes the last item heard, and the modality effect is accordingly reduced (Crowder & Morton, 1969). A similar, but modified, argument (see, e.g., Greene & Crowder, 1984) applies to "modality" and suffix effects achieved using nonacoustic stimuli, such as with mouthed or lip-read stimuli (e.g., Campbell & Dodd, 1980; Greene & Crowder, 1984; Nairne & Walters, 1983; Spoehr & Corin, 1978).

It is interesting that the modality effect can be restored if the participant is induced to group, or we prefer the term *stream* (cf. Bregman, 1990), the suffix in a different stream from the to-be-recalled stimulus. For instance, Morton, Crowder, and Prussin (1971) found that, for monaural list presentation, a contralateral suffix, or even a binaural suffix, was much less effective than an ipsilateral suffix in abolishing the modality effect. It is well known that perceived spatial location is a powerful cue to streaming; contralateral or binaural suffixes would tend therefore to be streamed separately from the list items. A large attenuation in the suffix effect was also found when the suffix was pronounced in a voice different from that in which the list items were pronounced, or even in the same voice but in a different pitch. In a streaming account, this corresponds to streaming by fundamental frequency, also known to be a powerful effect. Segregation of the list from the suffix can be achieved by presenting more than one suffix: The suffixes tend, retrospectively, to form a stream of their own (streaming by repetition). The ability of repeated suffixes to reduce the suffix effect has been shown in several studies (e.g., Crowder, 1978; Morton, 1976), though this ability is lost if the repeated suffixes are presented, on a given trial, in a voice different for each suffix (LeCompte & Watkins, 1995). In the latter situation, we presume, segregation by fundamental frequency overrides any potential streaming by repetition.

A small modification to our original hypothesis, whereby the last items heard in several recent streams are maintained independently, without mutual interference, can account qualitatively for all these results. Such a hypothesis might also account for the findings of Ayres, Jonides, Reitman, Egan, and Howard (1979) and Neath, Surprenant, and Crowder (1993). The latter, for example, presented participants with lists suffixed with the onomatopoeic word *baa*. The experimenters informed participants either that the suffix had a human (speech) origin or that it had been produced by a sheep (nonspeech). Larger suffix effects were obtained when the suffix was labeled speech as opposed to nonspeech. We speculate that the instructions led the former participants to perform schema-based streaming (Bregman, 1990), which segregated the suffix from the list items in the nonspeech condition, thus better preserving a record of the last stimulus item heard. Finally, our modified hypothesis might explain the fact that in objectively grouped lists, the modality effect is evident at the last position in each group (Frankish, 1974, 1985). The necessary assumption is that each group forms a separate stream.

Discussion

Theories of serial recall have been dominated by the view that order is encoded in terms of either position–item or item–item associations. The model presented here develops an alternative possibility: that order is represented as a primacy gradient of activation across list items. This simple idea, combined with decay and suppression of previously recalled items, gives rise to a model that gives a good account of a wide range of data on serial recall. Central to this account is an explanation of the form of the serial-position curve. Incorporation of decay automatically allows the model to explain the effects of both word length and list length. Furthermore, the commonly observed linear relationship between span and rate of articulation follows from the simple assumption that participants attempt to perform cumulative rehearsal between successive list items. Finally, the model can be extended to account for the effects of phonological similarity by adding a second stage of processing that is sensitive to the degree of phonological overlap between list items. In this regard, the model shares similarities with various models of speech production. The extended model can simulate data from Baddeley (1968) and Henson et al. (1996) that have proved beyond the scope of other quantitative models. In addition to these quantitative simulations, the model also provides a qualitative account of other data, including those related to the modality and suffix effects and to recall after short delays.

Comparison With the Perturbation Model

One of the most widely applied computational models of serial recall is the perturbation model of Estes (1972) and Lee and Estes (1977, 1981). In presenting our model, we have encountered the belief that the perturbation model has already accounted for many of the phenomena that our model seeks to explain. In this section, therefore, we discuss the similarities and differences between the models. Our task is complicated by the fact that we believe that the perturbation model first suggested by Estes differs fundamentally from that described by Lee and Estes and that the difference is widely unappreciated. The essential difference between the two models can be most simply expressed by noting that, whereas the Estes model is, like the primacy model, an order-based model, the Lee and Estes model is based on multilevel positional attributes and lacks an explicit recall mechanism. In the next section, we expand on this distinction before making comparisons with the primacy model.

The mechanism underlying the model described by Estes (1972) is clear. Representations of list items are connected to a contextual control element by "reverbatory loops" (p. 178). The item representations are cyclically reactivated, initially in the order specified by the input sequence. Over several reactivation cycles, cumulative perturbations in their timing can lead to interchanges in the order in which items are reactivated. Whenever recall is required, the items are simply read out in the order "determined by the current timing relationships" (Estes, 1972, p. 180). Like the primacy model, this version of the perturbation model has been implemented as a computer program that can simulate serial recall on a trial-by-trial basis. By contrast, in the model described in Lee and Estes (1981)

> information about the order of items is carried not by the order in which the items are currently being recycled in the rehearsal loop but rather by the current state of encoded information concerning attributes having to do with relative position among the events of a trial. (p. 164)

The model assumes a "multilevel associative structure" (p. 151) within which

> the memory representation of an individual item may be conceived as a vector of attribute information, including the current remembered position of the item within the sequence of the trials, its segment within the trial, and its position within a segment. The codes at each level are subject to perturbations, which lead at recall to transposition errors, that is, incorrect report of the trial, the segment within a trial, or the position within a segment in which the item occurred. (Lee & Estes, 1981, p. 151)

Note that it is now (unspecified) vectors of attribute information that are subject to independent perturbations of aspects of their content, as opposed to Estes' original conception of unitary item representations subject to perturbations in their timing. This change is fundamental. The Estes model and the primacy model describe what happens on an individual recall trial. Serial-position curves and transposition gradients are derived by running the model for many trials to simulate the outcome of a complete experiment. In contrast, the mathematical formulation of the Lee and Estes perturbation model generates transposition gradients (sometimes called simply *position gradients*) directly. No means is suggested by which the perturbed attribute vectors can be converted, on a given recall attempt, into an appropriate response sequence. In short, the later model is not a model of the process of serial recall.

Fundamental differences between the Estes (1972) model and the Lee and Estes (1981) model should perhaps not surprise us. The two models were developed to account for the results from experiments involving quite different tasks: The Estes model (like the primacy model) addressed the task of recalling span- or subspan-length sequences, either immediately or after a short delay. The Lee and Estes model addressed full and partial recall of 12-item, grouped lists under fast presentation conditions. Although the later model produced good simulations of the data to which it was applied, it cannot be so successfully applied to data from the ISR task. The chief problem with such retrospective application relates to assumptions regarding the timing of perturbations. Applications of the Lee and Estes perturbation model (e.g., Cunningham, Healy, Till, Fendrich, & Dimitry, 1993; Healy, Fendrich, Cunningham, & Till, 1987; Nairne, 1992) typically assume that there are no perturbations during input of stimulus items or during recall (the assumptions for grouped lists are slightly different, as we note later). This assumption is made so as to guarantee that the serial-position curves are completely symmetrical (Healy et al., 1987, p. 417). By contrast, serial-position curves of the type modeled here, that is, those describing performance on lists of a length approximately equal to span, reliably show marked asymmetry: The primacy portion extends almost throughout the list, with a small recency advantage being confined to the last item. Thus, none of the perturbation regimes proposed in the studies mentioned above could usefully apply to data from immediate serial-recall tasks; neither could they be adapted to do so while simultaneously preserving their ability to account for the data for which they were devised. The original perturbation regime, assumed by Lee and Estes to operate for grouped stimulus lists, for which perturbations within a given group are allowed during presentation and, at a reduced rate, during recall of other groups, is no more successful at simulating immediate serial recall: It predicts best performance on the last group, which, although consistent with their data, is quite contrary to the immediate serial-recall data.

Having made clear the distinction between the Estes (1972) and Lee and Estes (1981) models, we must make it clear that the primacy model bears a considerable family resemblance to the earlier of the two models. Where Estes considered early list items to be advanced in the order in which items are reactivated by parallel "reverbatory loops" (p. 178; the workings of which are not explored in detail), we consider such items to have a high activation. Where Estes suggested timing perturbations in the reactivation cycles, we suggest activation noise. Where Estes suggested cumulative perturbations, we suggest a constant activation noise acting on a decaying memory trace. The clear resemblances between the models follow from the fact that they are both order-based models, which address themselves directly to the immediate serial-recall task. The primacy model is more explicit in its precise mechanism, particularly in relation to omission errors, and also has a good deal of extra mechanism, reflecting the extended range of phenomena to which it is addressed. (We note that, to our knowledge, no perturbation model has ever been successfully applied to the modeling of the effects of phonological similarity, word length, or list length.) Nonetheless, even though the development of the primacy model was actually prompted by the work of Grossberg (1978, etc.), we believe it can, and should, also be seen as a direct descendant of the order-based model described earlier by Estes.

Grouping and Positional Intrusions

Our claim that serial recall does not involve reinstating specific positional cues appears to be challenged by data from experiments on grouping. From the point of view of assessing the role of positional information in recall, the most significant aspect of the recall of grouped lists is the pattern of transposition errors. In general, grouping gives rise to a pattern of errors in which transpositions between items in the same within-group position become approximately as common as one-apart transpositions (Frankish 1974; Lee & Estes, 1981; Ryan, 1969a, 1969b). It is hard to avoid the conclusion that within-group position is coded in some way. If within-group position is coded, then why not within-list position in general? In common with Wickelgren (1967), we suggest that these results could be achieved using positional information that is no more sophisticated than specifying the positions *beginning, middle,* and *end* of group. This restricted positional coding might explain why groups of three are optimal, because this is the maximum group size for which each within-group position is provided with a unique positional code (i.e., groups of four would be coded *beginning, middle, middle,* and *end*). Clearly, such crude positional associations would not, by themselves, be capable of supporting serial recall in general. The presence of such associations may interact, however, with the primacy gradient, this interaction becoming increasingly manifest in situations where, because of decay, the primacy-gradient activations are rather weak, as would be the case following presentation of long lists. Simulations of grouping data using the primacy model will be presented in a future article.

It has been suggested (e.g., Burgess & Hitch, 1992) that positional intrusions constitute good evidence for the use of positional coding during recall. These are errors whereby the erroneously recalled item comes from the same position in the previously presented list or the previous response. Note that, in order to be sure that intrusions really do originate from the previous list, one should best employ a paradigm in which the items sets used on consecutive trials are entirely distinct. Some data from Estes (1991) are relevant here: Estes found that items recalled on a given trial that were not in the stimulus for that trial, but were recalled in response to the previous trial, tended to be recalled at or near the same position as they had been in the previous trial's recall. There are, however, two things worth noting. Firstly, the data are derived by collapsing across three conditions, all of which are delayed-recall conditions. The delays range from six digits (3 s) to 18 items (9 s). These delays would likely be sufficient to eliminate most, if not all, of the contribution to serial recall of the phonological store (which we identify here with the primacy gradient). For instance, Bjork and Healy (1974) showed that the phonological similarity effect, often used as an index of the involvement of the phonological store, is drastically reduced after less than 4 s, and Estes (1972) himself stated that "the duration of the short-term process appears on the basis of considerable evidence to be no more than 2 to 3 seconds" (p. 179). It would thus be premature to propose a positional model for immediate serial recall, as opposed to delayed recall, based on these data. Indeed, it seems to us curious to use data concerning intertrial positional intrusions, where the intertrial duration is typically around 20 s, to motivate the design of a system that is apparently only effective over periods of less than 4 or 5 s.

Secondly, and perhaps more tellingly, Estes (1991) himself discussed the origin of these positional intertrial intrusions and attributed them to a "contextual component" (p. 168) that, he makes clear, is quite separate from the representation that "can support recall quite effectively at the retention intervals used in this study, but does not carry over from one trial to the next" (p. 168). Thus Estes, like us, assumes a dissociation between the representation used for recall and that responsible for positional intrusions. We note that this dissociation is supported both by Dillon and Thomas (1975) and by Bjork and Healy (1974), who showed that participants accord recalls that are intrusions from previous trials a much lower confidence rating than recalls of items that were indeed present in the most recent stimulus. Further support is provided by Tehan and Humphreys (1995), who showed that the short-term availability of phonemic codes, which they too identified with Baddeley's (1986) notion of the phonological store, can result in immunity to proactive interference in an immediate serial-recall test; this immunity disappears after 2 s of distractor-digit shadowing. These data are at least consistent with the view adopted here that intrusions and omissions occur only when the primacy-based activation falls below threshold. Under these circumstances, participants may resort to an alternative source of information, perhaps context based and positional, which itself is not sufficient to support good ordered recall.

Summary

We believe many of the features of the primacy model are vital if immediate serial recall is to be modeled successfully. Of these, the fill-in property, the provision of a two-stage mechanism for phonological confusions, and decay during unrehearsed delays are the most important and are those features that enable the primacy model to give a more complete account of the immediate serial-recall data than any of its competitors. The model is simple, and its performance can be modulated using a small number of free parameters, each of which has a clear functional interpretation. Accurate simulations, allied to the revival of the notion that immediate serial recall is perhaps mediated neither by item–item associations nor by position–item associations, lead us to believe that the primacy model represents significant progress in the modeling of immediate serial recall.

References

Anders, T. R., & Lillyquist, T. D. (1971). Retrieval time in forward and backward recall. *Psychonomic Science, 22,* 205–206.

Avons, S. E., Wright, K. L., & Pammer, K. (1994). The word-length effect in probed and serial recall. *Quarterly Journal of Experimental Psychology, 47A,* 207–231.

Ayres, T. J., Jonides, J., Reitman, J. S., Egan, J. C., & Howard, D. (1979). Differing suffix effects for the same physical stimulus. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 315–321.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18,* 362–365.

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

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

Baddeley, A. D., & Andrade, J. (1994). Reversing the word-length effect: A comment on Caplan, Rochon, and Waters. *Quarterly Journal of Experimental Psychology, 47A,* 1047–1054.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.

Baddeley, A. D., & Hull, A. (1979). Prefix and suffix effects: Do they have a common basis? *Journal of Verbal Learning and Verbal Behavior, 18,* 129–140.

Baddeley, A. D., & Lewis, V. (1984). When does rapid presentation enhance digit span? *Bulletin of the Psychonomic Society, 22,* 403–405.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36A,* 233–252.

Baddeley, A. D., & Scott, D. (1971). Short-term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology, 23,* 275–283.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575–589.

Bishop, D. V. M., & Robson, J. (1989). Unimpaired short-term memory and rhyme judgement in congenitally speechless individuals: Implications for the notion of articulatory coding. *Quarterly Journal of Experimental Psychology, 41A,* 123–140.

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

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Brown, G. D. A., & Hulme, C. (1995). Modelling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34,* 594–621.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21.

Burgess, N. (1995). A solvable connectionist model of immediate recall of ordered lists. In G. Tesauro, D. Touretzky, & T. K. Keen (Eds.), *Advances in Neural Information Processing Systems, 7,* (pp. 51–58). Cambridge, MA: MIT Press.

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

Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology, 32,* 85–99.

Caplan, D., Rochon, E., & Waters, G. S. (1992). Articulatory and phonological determinants of word-length effects in span tasks. *Quarterly Journal of Experimental Psychology, 45A,* 177–192.

Caplan, D., & Waters, G. S. (1994). Articulatory length and phonological similarity in span tasks: A reply to Baddeley and Andrade. *Quarterly Journal of Experimental Psychology, 47A,* 1055–1062.

Caprile, B., & Girosi, F. (1990). *A non-deterministic minimization algorithm* (AI Memo No. 104). Cambridge, MA: Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Cohen, M. A., & Grossberg, S. (1987). Masking fields: A massively parallel neural architecture for learning, recognizing and predicting multiple groupings of patterned data. *Applied Optics, 26,* 1866–1891.

Conrad, R. (1958). Accuracy of recall using keyset and telephone dial, and effect of a prefix digit. *Journal of Applied Psychology, 42,* 285–288.

Conrad, R. (1960). Very brief delay of immediate recall. *Quarterly Journal of Experimental Psychology, 12,* 45–47.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75–84.

Conrad, R. (1967). Interference or decay over short retention intervals? *Journal of Verbal Learning and Verbal Behavior, 6,* 49–54.

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

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31,* 668–684.

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.

Cowan, N., Wood, N. L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science, 5*(2), 103–106.

Crowder, R. G. (1967). Prefix effects in immediate memory. *Canadian Journal of Psychology, 21,* 450–461.

Crowder, R. G. (1978). Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review, 85,* 502–524.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics, 5,* 365–373.

Cunningham, T. F., Healy, A. F., Till, R. E., Fendrich, D. W., & Dimitry, C. Z. (1993). Is there really very rapid forgetting from primary memory? The role of expectancy and item importance in short-term recall. *Memory and Cognition, 21,* 671–688.

Dallett, K. M. (1964). Effects of a redundant prefix on immediate recall. *Journal of Experimental Psychology, 67,* 296–298.

Dallett, K. M. (1965). "Primary memory": The effects of redundancy upon digit repetition. *Psychonomic Science, 3,* 237–238.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27,* 124–142.

Dillon, R. F., & Thomas, H. (1975). The role of response confusion in proactive interference. *Journal of Verbal Learning and Verbal Behavior, 14,* 603–615.

Dosher, B. A., & Ma, J. (1998). Output loss or rehearsal loop? Output-time versus pronunciation-time limits in immediate recall for forgetting matched materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 316–335.

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

Drewnowski, A., & Murdock, B. B., Jr. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 319–332.

Ebenholtz, S. M. (1963). Serial learning: Position learning and sequential associations. *Journal of Experimental Psychology, 66,* 353–362.

Ellis, A. W. (1980). Errors in speech and short-term memory: The effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior, 19,* 624–634.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71,* 43–51.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). New York: Halsted Press.

Estes, W. K. (1973). Phonemic coding and rehearsal in short-term memory for letter strings. *Journal of Verbal Learning and Verbal Behavior, 12,* 360–372.

Estes, W. K. (1991). On types of item coding and sources of recall in short-term memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: In honor of Bennet B. Murdock* (pp. 155–173). Hillsdale, NJ: Erlbaum.

Ferguson, R. P., & Bray, N. W. (1976). Component processes of an overt rehearsal strategy in young children. *Journal of Experimental Child Psychology, 21,* 490–506.

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

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

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 177–220). London: Academic Press.

Granger, R., Myers, R., Whelpley, E., & Lynch, G. (1993). Intrinsic and emergent corticohippocampal computation. In M. Baudry & J. Davis (Eds.), *Long-term potentiation* (Vol. 2, pp. 371–392). Cambridge, MA: MIT Press.

Granger, R., Whitson, J., Larson, J., & Lynch, G. (1994). Non-Hebbian properties of long-term potentiation enable high-capacity encoding of temporal sequences. *Proceedings of the National Academy of Sciences, USA, 91,* 10104–10108.

Greene, R. L., & Crowder, R. G. (1984). Modality and suffix effects in the absence of auditory stimulation. *Journal of Verbal Learning and Verbal Behavior, 23,* 371–382.

Grossberg, S. (1977). A theory of human memory: Self-organization and performance of sensory-motor codes, maps and plans. In R. Roden & F. Snell (Eds.), *Progress in theoretical biology* (pp. 498–639). New York: Academic Press.

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., Fendrich, D. W., Cunningham, T. F., & Till, R. E. (1987). Effects of cuing on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 413–425.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A,* 80–115.

Hoosain, R. (1987). Language differences for pronunciation speed for numbers, digits and mathematical ability. *Psychologia, 30,* 34–38.

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.

Houston, J. P. (1965). Short-term retention of verbal units with equated degrees of learning. *Journal of Experimental Psychology, 70,* 75–78.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685–701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1–16.

Jahnke, J. C. (1975). Stimulus and response prefixes interfere differentially with short-term recall. *Journal of Experimental Psychology: Human Learning and Memory, 12,* 727–732.

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

LeCompte, D. C., & Watkins, M. J. (1995). Grouping in primary memory: The case of the compound suffix. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 96–102.

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

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lewandowsky, S., & Li, S. (1994). Memory for serial order revisited. *Psychological Review, 101,* 539–543.

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

Lowe, D. G., & Merikle, P. M. (1970). On the disruption of short-term memory by a response prefix. *Canadian Journal of Psychology, 24,* 169–177.

Lynch, G., & Granger, R. (1992). Variations in synaptic plasticity and types of memory in cortico-hippocampal networks. *Journal of Cognitive Neuroscience, 4,* 189–199.

Madigan, S. A. (1971). Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology, 87,* 294–296.

Mortenson, F. J., & Loess, H. (1964). Effect of very brief interpolated activity on short-term retention. *Perceptual and Motor Skills, 18,* 797–803.

Morton, J. (1976). Two mechanisms in the stimulus suffix effect. *Memory and Cognition, 4,* 144–149.

Morton, J., Crowder, R. G., & Prussin, H. A. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology, 91,* 169–190.

Murdock, B. B. (1961). The retention of individual items. *Journal of Experimental Psychology, 62,* 618–625.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21,* 263–276.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78,* 679–684.

Muter, P. (1980). Very rapid forgetting. *Memory and Cognition, 8,* 174–179.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory and Cognition, 18,* 251–269.

Nairne, J. S. (1992). The loss of positional certainty in long-term memory. *Psychological Science, 3,* 199–202.

Nairne, J. S., & Walters, V. L. (1983). Silent mouthing produces modality and suffix-like effects. *Journal of Verbal Learning and Verbal Behavior, 22,* 475–483.

Naveh-Benjamin, M., & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology, 38A,* 739–751.

Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin and Review, 2,* 429–441.

Neath, I., Surprenant, A. M., & Crowder, R. G. (1993). The context-dependant stimulus suffix effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 698–703.

Nigrin, A. L. (1993). *Neural networks for pattern recognition.* Cambridge, MA: MIT Press.

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

Norris, D., Page, M. P. A., & Baddeley, A. D. (1995). Connectionist modelling of short-term memory. *Language and Cognitive Processes, 10,* 407–409.

Page, M. P. A. (1993). *Modelling aspects of music perception using self-organizing neural networks.* Unpublished doctoral dissertation, University of Wales.

Page, M. P. A. (1994). Modelling the perception of musical sequences with self-organizing neural networks. *Connection Science, 6,* 223–246.

Page, M. P. A., & Norris, D. (1998). Modeling immediate serial recall with a localist implementation of the primacy model. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 227–255). Mahwah, NJ: Erlbaum.

Palmer, S. E., & Ornstein, P. A. (1971). Role of rehearsal strategy in serial probed recall. *Journal of Experimental Psychology, 81,* 60–66.

Peterson, L. R., & Johnson, S. T. (1971). Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior, 10,* 346–354.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Richman, H. B., & Simon, H. A. (1994). *EPAM simulations of short-term memory* (Complex Information Processing Working Paper No. 514). Pittsburgh, PA: Carnegie-Mellon University, Department of Psychology.

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

Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology, 21,* 148–155.

Shiffrin, R. M., & Cook, J. R. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behavior, 17,* 189–218.

Slamecka, N. J. (1967). Serial learning and order information. *Journal of Experimental Psychology, 74,* 62–66.

Spoehr, K. T., & Corin, W. J. (1978). The stimulus suffix effect as a memory coding phenomenon. *Memory and Cognition, 6,* 583–589.

Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition, 21,* 1–20.

Tehan, G., & Humphreys, M. S. (1995). Transient phonemic codes and immunity to proactive interference. *Memory and Cognition, 23,* 181–191.

Wickelgren, W. A. (1965a). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology, 70,* 102–108.

Wickelgren, W. A. (1965b). Short-term memory for phonemically similar lists. *American Journal of Psychology, 78,* 567–574.

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

Young, R. (1961). The stimulus in serial verbal learning. *American Journal of Psychology, 74,* 517–528.

Young, R. (1962). Tests of three hypotheses about the effective stimulus in serial learning. *Journal of Experimental Psychology, 63,* 307–313.

Young, R. (1968). Serial learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 122–148). Englewood Cliffs, NJ: Prentice Hall.

Appendix

An Activation-Based Mechanism for Generating a Primacy Gradient

A primacy gradient can result from an activation-based mechanism in the following way: Suppose a layer of item nodes, indexed by i, each of whose activations, x_i, varies such that

$$\frac{dx_i}{dt} = -Dx_i + (A - x_i)\cdot I_i, \quad (A1)$$

where D is the primacy model decay parameter and I_i is the input to a given node, which is set high (e.g., $I_i = 50$—the exact value is not critical) for a short period (e.g., 200 ms—again not critical) after presentation of the item corresponding to that node. A is given by

$$A = s\cdot(1 - \frac{n}{P}), \quad (A2)$$

where P is the model's peak parameter and n is the number of nodes active in the layer at the time of a given item's onset. Finally, s is the activation of a node that activates to a given level at the onset of a to-be-remembered list and decays thereafter with exponent D as above. The level to which this cell activates at list onset is arbitrary and can be set to P for definiteness.

These equations guarantee a decaying constant-step primacy gradient in response to the presentation of a list of items. The strength of the primacy gradient at the end of list presentation will depend on the time since list presentation began, as described in the main text.

Received May 19, 1995
Revision received April 16, 1998
Accepted April 20, 1998 ■

Editor's Note
Diversifying the Scope of Theoretical Notes in *Psychological Review*

Traditionally, Theoretical Notes in *Psychological Review*, with rare exceptions, have consisted of critiques of prior articles and replies to such critiques. As a matter of formal policy, the *Review* is now open to Theoretical Notes of multiple types, including, but not limited to, discussions of previously published articles, comments that apply to a class of theoretical models in a given domain, critiques and discussions of alternative theoretical approaches, and metatheoretical commentary on theory testing and related topics.

This initiative represents an effort to make *Psychological Review* the home for a broad range of theoretical commentary. There will be no change, however, in the *Review*'s policy of subjecting Theoretical Notes to a rigorous review for publication, nor will there be a change in the *Review*'s policies on critiques and replies (see the January 1996 issue). Theoretical Notes will continue to be distinguished from regular articles, not only by their appearing in the Theoretical Notes section of each issue, but also by wording such as "Critique of. . .," "Reply to. . .," "Comment on. . .," "Note on. . .," and so forth, in the titles of such articles.