

Phonological Similarity in Serial Recall: Constraints on Theories of Memory

Stephan Lewandowsky

University of Western Australia

and

Simon Farrell

University of Bristol

Running head: Phonological Similarity

Address correspondence to: Stephan Lewandowsky

School of Psychology

University of Western Australia

Crawley, W.A. 6009

lewan@psy.uwa.edu.au

Abstract

In short-term serial recall, similar-sounding items are remembered more poorly than items that do not sound alike. When lists mix similar and dissimilar items, performance on the dissimilar items is of considerable theoretical interest. S. Farrell and S. Lewandowsky (2003) recently showed that if guessing strategies are controlled, dissimilar items on mixed lists are recalled more accurately than on pure dissimilar lists, a finding that challenges several current theories of serial recall. This article presents two experiments that extend the generality of the mixed-list advantage for dissimilar items and then applies three theories of memory—the primacy model, SIMPLE, and SOB—to the data. The simulations show that the data are best explained by the SOB theory (S. Farrell, 2006; S. Farrell & S. Lewandowsky, 2002) which, unlike most other current theories, posits that similarity has an effect at the time of encoding.

Phonological Similarity in Serial Recall: Constraints on Theories of Memory

The phonological similarity effect in short-term serial recall refers to the well-replicated finding that lists composed of similar-sounding items are less accurately recalled in the correct order than lists in which items do not sound alike (e.g., Baddeley, 1966, 1968; Conrad, 1964; Henson, Norris, Page, & Baddeley, 1996; Wickelgren, 1965a, b). This effect is of considerable generality, occurring with consonants (Baddeley, 1968) as well as with words (Baddeley, 1966; Coltheart, 1993; Henry, 1991). The phonological similarity effect is typically considered a benchmark result of short-term serial recall that must be accommodated by relevant theories (e.g., Brown, Preece, & Hulme, 2000; Nairne & Kelley, 1999).

The phonological similarity effect also occurs when phonologically similar (e.g., B, P, T) and dissimilar (e.g., K, Q, R) items are mixed together on study lists. Mixed lists are particularly diagnostic because the level of recall of the dissimilar items can differentiate between competing theories of memory. Initial reports (e.g., Baddeley, 1968; Bjork & Healy, 1974; Henson et al., 1996) suggested that dissimilar items on mixed lists are recalled with the same accuracy as items on pure dissimilar lists. This apparent absence of an effect of intermixing similar items on the recall of dissimilar items compelled several theorists to propose that serial recall involves two independent stages of processing, with order errors occurring between positional tokens in a primary stage, and with similarity-based confusions occurring in a separate secondary stage that operates only on similar items (e.g., Burgess & Hitch, 1999; Henson, 1998; Page & Norris, 1998a, b).

More recently, Farrell and Lewandowsky (2003; see also Farrell, 2006) showed that when guessing strategies are controlled—either by presenting all list items for re-ordering at test or by equalizing total ensemble size across list types—dissimilar items on mixed lists were in fact better recalled than their counterparts on pure dissimilar lists (we refer to this finding as a

“mixed-list advantage” from here on). Farrell and Lewandowsky (2003) suggested that their results challenged two-stage theories and proposed that models in which similarity affects encoding processes, such as the feature model (e.g., Neath, 2000) or SOB (Farrell & Lewandowsky, 2002), provided a more suitable explanation. Farrell (2006) demonstrated that the SOB model accounted for the mixed-list advantage at a quantitative level, whereas a two-stage model failed to provide an adequate account. However, the work by Farrell and Lewandowsky (2003) and Farrell (2006) was limited in two ways: First, at an empirical level, the mixed-list advantage could be explained by appealing to encoding or retrieval strategies. Second, although the mixed-list advantage promises to differentiate between two classes of theory, its theoretical implications still await thorough examination. This article addresses both of those limitations.

We first report two mixed-list experiments that extend the findings of Farrell and Lewandowsky (2003) using a procedure that reduces the possible role of encoding or retrieval strategies. We then compare three theories of memory by simulation and examine their ability to handle the mixed-list advantage for dissimilar items. The simulations show that the two-stage primacy model (Page & Norris, 1998a, b) cannot accommodate the mixed-list advantage. The SIMPLE temporal distinctiveness theory (Brown, Neath, & Chater, 2002) produces a mixed-list advantage with some, but not all, parameter values. The SOB theory of Farrell and Lewandowsky (2002) and Farrell (2006), by contrast, reliably produces a mixed-list advantage and overall provides the best quantitative account of the data. The simulation results suggest that similarity effects are best modelled by assuming that similarity affects encoding into memory rather than contributing only to confusions at retrieval.

MIXED LIST PHONOLOGICAL SIMILARITY IN SERIAL RECALL:
THE EMPIRICAL PATTERN

The use of lists that contain a mixture of items of different types has been a highly diagnostic tool in several areas of memory research. For example, in recognition memory, lists containing a mixture of strong and weak items (e.g., items that are presented for 1 s and 2 s, respectively) have been central to arguments about the relative merits of various computational models (e.g., Murdock & Kahana, 1993a, b; Shiffrin, Ratcliff, Murnane, & Nobel, 1993). Although adjudication between those models has been based on a plethora of data, the influence of the—mixed-list-based—list strength effect has been remarkable. In serial recall, recent work on the word length effect has likewise exploited the diagnosticity of mixed lists (Hulme, Neath, Stuart, Shostak, Surprenant, & Brown, 2006). One reason for the widespread utility of mixed lists is that relevant models inevitably handle the benchmark differences between pure lists of the different item types under consideration: For example, models of recognition memory predict performance to be better on pure strong than pure weak lists; models of short-term memory can accommodate the ubiquitous word length effect and they also universally predict that lists of similar items will be recalled worse than lists of dissimilar items. Models may differ, however, in their ability to handle the results from mixed lists.

Concerning phonological similarity, Baddeley (1968, Experiment 5) reported the first experiment in which items within a list were either mutually dissimilar (e.g., M K and R) or highly similar (e.g., V, D, and T), with the two types being interleaved to form a list such as M V K D R T. This mixed list is commonly referred to by its abstract code as DSDSDS, where S and D represent, respectively, a similar and dissimilar item. Baddeley (1968) found that the accuracy with which D items were recalled from pure lists (DDDDDD) was identical to that of D items in the corresponding serial positions on mixed lists (e.g., SDSDS). Henson et al. (1996) replicated

the result using a more refined method and analysis. The absence of a mixed-list advantage for dissimilar items was interpreted as evidence against chaining models (e.g., Lewandowsky & Murdock, 1989; see Henson et al., 1996, for further discussion) and became a benchmark result for theorizing during the 1990's (see Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998; Henson et al., 1996; Page & Norris, 1998a).

Notwithstanding its acceptance for an extended period, the absence of a mixed-list advantage is quite counter-intuitive. Consider the extreme case of a single dissimilar item being embedded in a similar-sounding list (e.g., the letter *X* in the list B D G X T P); on the basis of the ubiquitous isolation effect—often called the “von Restorff effect” after its initial investigator—the dissimilar item should be recalled more accurately in this case than when it is surrounded by other dissimilar items. Indeed, the presence of other isolation effects in serial memory tasks (e.g., Bone & Goulet, 1968; Cimbalò, Nowak, & Soderstrom, 1981; Kelley & Nairne, 2001; Lippman, 1980) might lead one to question why there was no facilitation for dissimilar items on the mixed lists in the studies by Baddeley (1968) and Henson et al. (1996).

Farrell and Lewandowsky (2003) pursued this question and found that recall of dissimilar items is unaffected by the presence of similar items only when people can trade off order and item errors. Hence, even when the pattern of correct responses gave the appearance that recall of dissimilar items did not differ between list types (i.e., pure D vs. mixed), the underlying transposition rates indicated that order memory was better for dissimilar items on mixed lists than on pure lists. Performance appeared equal only because of compensatory differences in intrusions and omissions (Experiment 1). When item errors were controlled by use of a reconstruction task (Experiment 2) or were equalized across list types by equalizing vocabulary size (Experiment 3), a mixed-list advantage for dissimilar items was observed even at the level of correct responses. To give an impression of the consistency of those results, Figure 1 summarizes the data from all

experiments of Farrell and Lewandowsky (2003) by plotting the transposition error rates for dissimilar items on mixed lists (mixed D) against transposition rates for items in corresponding serial positions on pure lists (pure D). Contrary to the early results by Baddeley (1968) and Henson et al. (1996), nearly all data points deviated systematically and considerably from the diagonal (which would represent equal performance on both list types), such that there were always more transpositions for an item in a pure dissimilar list than in a mixed list.

Although these results are consistent and replicable (e.g., Farrell, 2006), two plausible objections can be raised against the procedure of Farrell and Lewandowsky (2003). First, in their studies the various list types were presented contiguously in blocks of 40 trials of the same type. Although this followed precedent (Henson et al., 1996), the blocked presentation regime may have induced people to develop strategies that were specific to each list type, such as encoding of the mixed lists into two separate “streams” of similar and dissimilar items. Farrell and Lewandowsky (2003) sought to address this problem by showing that the mixed-list advantage did not differ between early and late trials within a block; nonetheless, the development of strategies cannot be completely ruled out. In recognition of this issue, Farrell (2006) presented lists in a random order, and still obtained a sizeable mixed-list advantage. However, the mixed lists of Farrell (2006) contained only a single dissimilar item, which gives rise to a second concern.

In the studies by Farrell and Lewandowsky (2003) and Farrell (2006) some lists contained a single dissimilar item that was surrounded by similar items (e.g., B D G X T P). In consequence, the overall experimental sequence contained more similar than dissimilar items, thus creating an imbalance in potential proactive inhibition or “token dose” (Bridges & Jones, 1996) between item types. We now report two experiments that resolved both potential objections to the earlier studies.

MIXED-LIST EXPERIMENTS

Experiment 1

The first experiment contained alternating lists (SDSDSD and DSDSDS), lists with a single dissimilar item (SSSDSS and SSSSSD; called isolate lists), and pure dissimilar lists. In order to reduce the imbalance between the number of presentations of similar and dissimilar items, and because emphasis here was on the comparison between mixed and pure lists for dissimilar items, Experiment 1 did not include any pure similar lists (SSSSSS).

Moreover, like the study of Farrell (2006) and unlike the studies by Farrell and Lewandowsky (2003), list types in Experiment 1 were intermingled in random order to prevent the build-up of specific encoding strategies. There was, however, one important exception to the randomization of list types: The pure dissimilar lists (DDDDDD) were withheld until the later stages of the experiment, such that the last 20% of trials all involved the pure dissimilar lists and the first 80% of trials involved a random ordering of the other 4 list types (SDSDSD, DSDSDS, SSSDSS, and SSSSSD). Withholding of the pure lists is required to keep ensemble size equal across list types: Farrell and Lewandowsky (2003) showed that guessing strategies are a critical factor in mixed-list performance, and because mixed lists are necessarily drawn from two pools (i.e., some number of similar items plus a set of dissimilar items), guessing opportunities for pure dissimilar lists can be equalized only by drawing items from a larger pool (i.e., double the size of the dissimilar pool used for mixed lists). This requirement, in turn, mandates that the full dissimilar ensemble not be revealed until the end of the experiment. Farrell and Lewandowsky showed that this manipulation controlled guessing strategies.¹

Method

Participants and Design

Thirty-three members of the campus community at the University of Western Australia participated voluntarily and were remunerated at the rate of \$10/hr.

The single experimental variable, list type, was manipulated within participants. There were 5 different types of 6-item lists: Two isolate lists containing a single dissimilar item in position 4 or 6 (SSSDSS and SSSSSD, respectively), two alternating lists (SDSDSD and DSDSDS), and the pure dissimilar control list (DDDDDD). 24 lists of each type were constructed for each participant.

Materials

Lists were constructed by randomly sampling letters from three 5-letter ensembles. All similar stimuli were drawn from the ensemble B, T, D, G, P. Dissimilar letters were drawn from one of two dissimilar ensembles consisting of the 5 letters H, K, M, Q, R, and X, L, Z, W, Y, respectively. (The letter “Z” in Australian usage does not rhyme with “B” or “T” but is pronounced “zed.”)

All mixed lists were constructed from the similar ensemble and one of the dissimilar ensembles, with the identity of the dissimilar ensemble counterbalanced across participants. For example, one participant might receive an SSSDSS list composed of G, P, D, K, T, B, where the K could be replaced by any of M, Q, R, and H on other trials for the same list type. Another participant would receive lists of the kind G, P, D, X, T, B, where the X could be any of L, Z, W, and Y across trials for that list type. The other mixed lists (SSSSSD, DSDSDS, and SDSDSD) followed the same structure, thus ensuring that all mixed lists were drawn from a total vocabulary of 10 potential stimuli for each participant. The pure dissimilar lists (DDDDDD) were sampled

from both dissimilar ensembles for all participants, thus equalizing ensemble size for mixed and pure lists. Familiar sequences or acronyms (e.g., TV, BHP) were disallowed.

Because serial recall performance can be affected by the predictability of list items (cf. Henson et al., 1996), we computed the average predictability of stimuli within list type using the logarithms of bigram occurrence in the corpus of Solso and Juel (1980). For example, for the DDDDDD list, this measure consisted of the average bigram frequency across all 90 possible combinations of H, K, M, Q, R, X, L, Z, W, and Y (excluding repetitions). Table 1 shows that predictabilities differed somewhat between the various lists, with the largest difference being around .40. This largest difference was marginally smaller than the largest difference between list types in Henson et al.'s (1996) Experiments 2 and 3. (Their means ranged from .70 to 1.14, so proportionally their spread was considerably larger than ours.) Moreover, our largest difference was less than 25% the magnitude of the largest difference in Henson et al.'s Experiment 1, suggesting that list predictability was adequately controlled in the present study.

All mixed list types (i.e., SSSDSS, SSSSSD, SDS DSD, and DSDSDS) were presented in random order during the first 96 trials, whereas the last 24 trials were composed of pure dissimilar lists (DDDDDD). This ensured that the second dissimilar ensemble that was needed for the DDDDDD lists did not become part of the experimental vocabulary during the earlier mixed-list trials.

Procedure

Participants were informed that they would be presented with lists of 6 letters not containing any repeated items. For each list, the upper-case letters appeared one by one in the middle of the screen, each item being overwritten by its successor. Letters appeared for 400ms, with a 100ms inter-stimulus-interval (ISI). Participants were instructed to read the letters in silence.

Participants recalled the list immediately after the last item was presented by keyboard entry. Each recalled letter appeared on the screen, replacing the previously recalled item. Participants were not allowed to correct any mistakes. Participants were instructed to type the first letter that came to mind if they were unsure, but they could record an omission by pressing the space bar if necessary. Participants were instructed to repeat items if, later in the list, they felt a letter had appeared there rather than at the earlier position of its first recall. A self-paced break was inserted after every 24 trials.

Results and Discussion

Responses were considered correct only when an item was recalled in its correct serial position. The overall level of accuracy of two participants (.38 and .32) fell so far below the mean of .68 ($s = .15$) that they were eliminated from the analyses. None of the conclusions change if data from all participants are included.²

Serial Position Analysis

A necessary consequence of the present design is that the pure dissimilar lists benefit from experimental practice, and conversely, that the mixed lists presented early in the experiment suffer from lack of practice. Examination of overall performance levels during the first 4 blocks of 24 trials (all of which involved a random ordering of the same list types) confirmed that performance increased from .64 to .72 from the first to the fourth block, with much of that improvement having been achieved by the second block (.69). In order to provide an unconfounded comparison between the mixed lists and the pure lists, all trials from the first block were therefore considered to be practice trials and were excluded from all analyses. Note that notwithstanding the elimination of the first block, practice effects continue to militate against the detection of a mixed-list advantage; the design of both present experiments is therefore biased

against the hypothesis of interest. None of the results reported next change qualitatively if all trials are included.

The serial position curves for all list types are shown in the left-hand panel of Figure 2. Although the overall pattern is complex, the figure clearly reveals a zig-zag pattern for mixed lists, with dissimilar items being recalled better than similar items. In addition, the data suggest that dissimilar items on mixed lists were more accurately recalled than on pure lists, although this effect is more readily visible if the data are aggregated across the various mixed lists. Because the comparison of greatest interest involved the alternating lists (DSDSDS and SDSDS), a composite serial position curve was formed by combining responses to dissimilar items on the two alternating lists. The right-hand panel of Figure 2 shows this composite curve in comparison to the pure dissimilar list.³ The panel confirms that there was a consistent but small advantage for dissimilar items presented in alternating lists compared to pure lists.⁴ Accordingly, the 2×6 (List Type: alternating vs. pure \times Serial Position) within-subjects ANOVA revealed strong main effects of serial position, $F(5, 150) = 38.65$, $MSe = .025$, $p < .0001$, and list type, $F(1, 30) = 10.16$, $MSe = .016$, $p < .005$, but no interaction between the two variables, $F(5, 150) = 1.55$, $p > .10$. We defer discussion of the magnitude of the mixed-list advantage until after all the data have been presented.

Error Analysis

Omissions were found to be virtually non-existent, and we therefore focused on extra-list intrusions to examine potential guessing strategies. The intrusion rates for the various mixed lists ranged from .09 to .11 (average .10), which was identical to the rate of .10 for the pure dissimilar lists. These intrusion rates were very close to those observed by Farrell and Lewandowsky (2003, Experiment 3). The fact that intrusion rates were nearly equal across all list types suggests that

differential opportunities for guessing could not have contributed to the observed mixed-list advantage.

In addition, to examine whether the introduction of new items late in the experiment might have affected performance on the pure dissimilar lists, we classified the intrusions during pure D trials into “old” intrusions (i.e., an extra-list item from the initial dissimilar set); “new” intrusions (i.e., from the second dissimilar set introduced for the pure D trials), and “other” intrusions (i.e., any extra-list response belonging to neither of these categories). The proportion of overall responses that fell into those three categories was .033, .037, and .031, respectively. Finally, for the pure dissimilar lists, we compared recall of “old” items (i.e., those used in the initial mixed-list trials) to recall of “new” list items (i.e., those withheld until the pure D lists). A 6×2 (serial position \times item status: old vs. new) within-subjects ANOVA uncovered no effects involving item status, with both F 's < 1 . The fact that performance did not differ between old and new items in the final block and the fact that intrusion rates were equal for both classes of items suggests that people's performance on the pure D trials was not affected by the introduction of additional items late in the experiment.

Experiment 2

The first study sought to control encoding strategies by presenting the crucial mixed lists in random order. However, across all lists, similar items in the first experiment still outnumbered dissimilar items, thus giving rise to an imbalance in token-dose between the two classes of items.

Experiment 2 therefore used only alternating lists (SDSDSD and DSDSDS) in addition to the pure dissimilar control lists (DDDDDD; again withheld until the end of the session for the reasons cited earlier). In consequence, across all mixed lists, the number of similar and dissimilar items was equal, and across the entire experiment, fewer similar than dissimilar items were

shown, thus reversing the token imbalance compared to Experiment 1. In addition, participants in Experiment 2 were asked to read the stimuli aloud to maximize phonological encoding of items.

Method

Participants and design. Forty-six members of the University of Western Australia campus community participated voluntarily in exchange for remuneration of A\$10.

Stimuli and procedure. Lists were sampled as in Experiment 1 except that there were now 40 lists of each type. Accordingly, the first 80 trials involved the two alternating lists in random order and the last 40 trials involved the pure dissimilar lists.

Unlike Experiment 1, participants read the lists aloud. To facilitate reading aloud, presentation duration of each letter was 900ms, with a 100ms inter-stimulus-interval (ISI). A self-paced break was inserted every 30 trials. In all other respects, the procedure was identical to Experiment 1.

Results and Discussion

Responses were considered correct only when an item was recalled in its correct serial position. The overall level of accuracy of one participant (.19) constituted a clear outlier and this individual was removed from all analyses.

Serial Position Analysis

Examination of performance across blocks of 10 trials revealed a small practice effect during the first three blocks (performance increased from .73 to .76), a large increase from there to block 4 (.81), and a small practice effect from there on through to the last block of mixed-list trials (.84). Accordingly, the first three blocks (i.e., 30 trials) were considered practice trials and were not included in the remaining analyses.

The serial position curves for all list types are shown in the left-hand panel of Figure 3. To facilitate comparison, the composite serial position curve for dissimilar items on mixed lists is again shown together with the pure dissimilar list in the center panel of the figure. As in Experiment 1, there was a clear mixed-list advantage at all but the terminal serial positions, where performance may have been sufficiently close to the ceiling to prevent the effect from emerging. The corresponding 2×6 (List Type: alternating vs. pure \times Serial Position) within-subjects ANOVA revealed main effects of serial position, $F(5, 220) = 20.72$, $MSe = .007$, $p < .0001$, and list type, $F(1, 44) = 5.67$, $MSe = .015$, $p < .05$, but no interaction between the two variables, $F(5, 220) = 1.51$, $p > .10$.

Error Analysis

Omissions were again very infrequent. Intrusion rates were .057 and .054 in the two mixed conditions (SDSDSD and DSDSDS, respectively), compared to .039 for the pure dissimilar lists. Although these intrusion rates were somewhat lower in absolute terms than in Experiment 1 and in Farrell and Lewandowsky's (2003) Experiment 3, the percentage point difference between the pure and mixed lists was exactly the same as that reported by Farrell and Lewandowsky.

We performed a number of additional analyses to examine whether the introduction of new items late in the experiment might have affected performance on pure D lists. We first computed performance across blocks of 10 trials for the pure D lists and found it to be virtually invariant (.84, .85, .85, .84, and .85, respectively). The constancy of performance across blocks provides a first indication that people were not stymied by the introduction of the second dissimilar item set. As in Experiment 1, we compared recall of "old" D items to recall of "new" D items. A $4 \times 6 \times 2$ (block \times serial position \times item status: old vs. new) within-subjects ANOVA uncovered no effects involving either block or item status, with the largest effect (for serial position \times item status) falling short of significance, $F(5, 220) = 1.42$, $MSe = .03$, $p > .10$.

Finally, we examined the type of intrusions committed during the pure D trials by classifying extra-list responses as either an “old” or “new” intrusion (i.e., whether an extra-list response belonged to the initial or the second dissimilar set, respectively), or as an “other” intrusion. The proportion of overall responses that fell into those three categories was .013, .014, and .012, respectively. When intrusions were further broken down by whether the target item (i.e., the correct response for a given output position) belonged to the old or new dissimilar set, the proportions were, respectively, .014, .014, and .012 for new targets and .012, .014, and .011 for old targets.

Taken together, there is no indication that the introduction of the extended dissimilar set late in the experiment affected performance. On the contrary, the data suggest that performance by the end of the experiment had become stable and that people did not treat the new items any differently from the old items. In particular, there is no indication that introduction of the items necessitated any compensatory learning of the new set, thus allaying fears that the mixed-list advantage might have arisen from unduly deflated performance on the pure dissimilar lists.

Conclusions from Experiments

We conclude that in properly controlled experiments, a mixed-list advantage occurs for dissimilar items in serial recall. The effect occurs even in situations in which list types are randomly intermingled to the extent possible (Experiments 1 and 2) and it occurs irrespective of whether the experiment overall contains more similar or more dissimilar items (Experiment 2). Moreover, the effect occurred in the present studies notwithstanding the fact that the design was biased against its detection by withholding the pure dissimilar comparison list until participants were highly practiced.

Farrell (2006) has additionally shown that the mixed-list advantage endures over a filled delay between study and test, although his studies used only single-isolate lists. Altogether, the

mixed-list advantage has now been shown in 6 experiments involving 158 participants. The effect has been shown with a reconstruction task when pure dissimilar lists were used throughout the experiment (Farrell, 2006; Farrell & Lewandowsky, 2003) and with standard serial recall when guessing was controlled by withholding the pure dissimilar list until the end of the experiment (Experiments 1 and 2 here; Farrell & Lewandowsky, 2003).

By contrast, the earlier claims that dissimilar items do not benefit from being on mixed lists (Baddeley, 1968; Henson et al., 1996) relied on a null effect and were based on a procedure that demonstrably did not control guessing (see Experiments 1 vs. 3 in Farrell & Lewandowsky, 2003).

As we show next, the mixed-list advantage is not only of considerable generality but also turns out to be theoretically highly diagnostic. In particular, we show that although the effect was numerically small in the present experiments, its size was sufficient to place meaningful constraints on the various theories.

MIXED-LIST EFFECTS: THEORETICAL ACCOUNTS

The reliable existence of a mixed-list advantage challenges models in which all effects of similarity are relegated to a separate confusion stage (e.g., Page & Norris, 1998a, b; Burgess & Hitch, 1999). By the same token, the mixed-list advantage may be more readily accommodated by single-stage theories which assume that similarity is one of several factors affecting encoding (e.g., Farrell & Lewandowsky, 2002). We now present various models of the two types before comparing their predictions by simulation.

Two-Stage Models

The Primacy Model

In the primacy model (Page & Norris, 1998a, b), items are represented by the activation of item nodes. The activation level is assumed to be maximal for the first item, and is assumed to decrease across successive items, thus giving rise to a “primacy gradient” of activations. Retrieval occurs by a process known as competitive cueing, which at any given output position will select the most strongly activated item for recall. On the additional assumption that each recalled item is suppressed and thus unavailable for further report, a primacy gradient in conjunction with (noisy) competitive cueing will necessarily yield forward serial recall.

The primacy model handles similarity effects by the addition of an extra “confusion” stage, where confusion of items occurs based solely on their phonological similarity. Thus, the first stage of the model is thought to store items and their order without regard to their phonological similarity. The output of the first stage, produced by competitive cueing, is passed to the second stage, where phonological confusions are assumed to take place. Critically, these confusions only occur between similar items, such that items dissimilar to other list items will pass through the second stage unaffected. Because dissimilar items bypass this stage irrespective of list composition, and because order among items is exclusively maintained by the similarity-insensitive first stage, it follows that the primacy model cannot predict a mixed-list advantage for dissimilar items.

The tight and seemingly inseparable link between the empirical absence of a mixed-list advantage and the theoretical need for separate stages was underscored by Page and Norris (1998b), who reported that explorations of several single-stage models were unsuccessful and that “... the data alone appear to force us to accept a two-stage model...” (p. 243; see also Page & Norris, 1998a, p. 774). By implication, the discovery of a reliable mixed-list advantage for

dissimilar items may necessitate a re-evaluation of the primacy model; below, we report simulations that confirm the need for theoretical revision.

The Start-End Model and the Burgess-and-Hitch Model

Henson's (1998) Start-End Model (SEM) likewise assumes that similarity effects arise at a separate stage in which phonological confusions take place after normal retrieval operation of the model (see Henson, 1998, for details). The Burgess and Hitch (1999) model also holds that item selection, which is driven by a representation of encoded context, is unaffected by phonological similarity. (This stage loosely corresponds to the first stage in the Primacy model and SEM.)

Although the primacy model, SEM, and the model by Burgess and Hitch use very different mechanisms to represent order, their predictions concerning the effects of phonological similarity are tightly linked. Because these models share the two-stage assumption, they are committed to predicting that a mixed-list advantage should be absent (see Figure 12 of Burgess & Hitch, 1999, and Figure 13 of Henson, 1998). From here on we therefore focus only on the primacy model; conclusions regarding mixed-list effects in the primacy model likely extend to SEM or the Burgess and Hitch model by implication.

Phonological Representations in Two-Stage Models

The predicted absence of a mixed-list advantage in two-stage models is tied to strong assumptions about the phonological similarity structure among items. Specifically, to predict null mixed-list effects, the primacy model, SEM, and the Burgess and Hitch model have thus far assumed that dissimilar items bear no resemblance either to each other or to the similar items, thus ensuring that dissimilar items bypass the phonological confusion stage. However, this simplifying assumption is unrealistic: Using a multidimensional scaling solution for acoustic confusions, Farrell (2006) showed that the average pairwise similarity between dissimilar items is higher than that between similar and dissimilar items. When these more realistic assumptions

were incorporated into a two-stage model, Farrell (2006) found that even the SEM predicted a small mixed-list advantage, although as would be expected on the basis of the preceding analysis, its magnitude fell short of the magnitude of that observed empirically. Given the demonstrated importance of incorporating realistic assumptions about the phonological similarity structure among stimuli, all simulations reported below constructed lists on the basis of multidimensional scaling of acoustic confusions.

Single-Stage Models

A Computational Distinctiveness Account (SIMPLE)

A computational instantiation of temporal distinctiveness, known as SIMPLE (Scale Invariant Memory, Perception, and LEarning), was recently provided by Brown et al. (2002; see also Brown & Chater, 2001). SIMPLE makes three core assumptions: First, items are represented in terms of their position within a multi-dimensional psychological space. In all cases, one of those dimensions represents the (logarithmically transformed) time that has elapsed since presentation of an item. Additional dimension(s) are used to capture a variety of stimulus properties; in the present case, phonological similarity. Second, the similarity between any two items in memory is a declining function of the distance between them in that psychological space. Third, the probability of recalling a given item is inversely proportional to that item's summed similarity to all other potentially recallable items. Accordingly, items that are isolated from their neighbors—either in time or similarity space—are recalled particularly well.

It follows that the predictions of SIMPLE for mixed lists may depend on the particular representation of the mixed lists, and the particular distance relationships between neighboring items. A mixed-list advantage could be observed if a dissimilar item is more distant, in similarity space, from a “similar” neighbor (where “similar” refers to the similarity between that neighbor and other list items) than a dissimilar neighbor. We show below that the predictions of SIMPLE

are indeed sensitive to the particular experimental circumstances and the model's parameter values.

Serial-Order in a Box Model (SOB)

The SOB model of Farrell and Lewandowsky (2002), like the primacy model (Page & Norris, 1998a), relies on a primacy gradient in conjunction with competitive cueing to represent order among items. Unlike the primacy model, SOB endogenously generates the gradient by comparing incoming list items to the contents of memory (viz., a composite matrix of connection weights that represents all previous items). Encoding strength is a direct function of the novelty of incoming items. In the generalized case, when items can be similar to each other (Farrell, 2006), the encoding strengths of successive items decrease across serial positions. The extent of that decrease is greater for similar items than dissimilar items.

To understand how this similarity-sensitive encoding predicts that inter-item similarity will affect the ordering of *all* items, consider the expected gradients of encoding strengths for two three-item lists: one consisting only of dissimilar items (i.e., DDD) and one that has a dissimilar item placed between two similar items (SDS). The first item is stored with the same strength on both lists, as there are no previous items which could affect its novelty. The second item is also stored with the same strength in both cases, as it is equally novel with respect to the first item. However, the third item on the SDS list receives a smaller weight than its sibling on the DDD list, because its similarity to the first item is detected during the pre-encoding comparison to the contents of memory.

In combination with competitive cueing at retrieval, this encoding scheme predicts that the isolated dissimilar item will be recalled more accurately than its counterpart on the DDD list. This is because once the first item is recalled and suppressed, and thus removed from the competition of to-be-recalled items at the second output position, it is the relative difference of encoding

strengths between the second and third item that determines recall performance. Because the third item on the SDS list, having been encoded with less strength, will not offer as much competition as its counterpart on the DDD list, the isolated dissimilar item is predicted to be recalled more accurately. Support for this prediction was provided by Farrell and Lewandowsky (2003), Farrell (2006) and by the preceding experiments.

Farrell (2006) provided a first demonstration of SOB's predictions of a mixed-list advantage. Farrell applied SOB and SEM to mixed-list effects, with an emphasis on the effects of a filled delay in between study and recall. Farrell found that SOB provided a good quantitative account of the data under both delay conditions, whereas SEM, as described above, was unable to produce a mixed-list advantage of the size observed in the experiments. Below we carry out a more systematic exploration of SOB's predictions of a mixed-list advantage, and examine its quantitative account for a wider range of mixed-list findings.

SIMULATIONS

Three theories were compared by simulation to examine their predictions for mixed-list phonological similarity effects: The primacy model, SIMPLE, and SOB. To ensure a fair comparison between theories, 4 free parameters were estimated for each model. The simulations were designed with particular emphasis on capturing the true similarity structure among experimental stimuli. The models were fit to the data of the present experiments and to Experiment 3 of Farrell and Lewandowsky (2003). The data of the latter study, which for ease of exposition we refer to as Experiment 3 from here on, are shown in the right-hand panel of Figure 3. We modeled Experiment 3 because its isolate lists contained the single dissimilar item in different serial positions from Experiment 1, thus providing additional constraints for the models.

Similarity Representation and Global Simulation Parameters

Input for all simulations was constructed using a multi-dimensional scaling (MDS) solution that captured the similarity relations among the experimental stimuli. Hull (1973) presented a complete auditory confusion matrix for letters and digits based on roughly 10,000 listening errors committed by 135 participants. The confusion matrix for the 25 letters (“O” was omitted) was analyzed by MDS using a Euclidean distance model. The badness-of-fit values for scaling solutions in one through five dimensions, respectively, were .32, .17, .12, .09, and .08 (using Kruskal’s, 1964, stress measure). Accordingly, the three-dimensional solution was chosen as the most appropriate stimulus representation for the simulations, with the coordinates shown in Table 2. The MDS coordinates were used to construct the input for all models, with the details of the various implementations provided below. Because the simulations (unlike the experiments) included pure similar lists to provide a broader picture of the models’ predictions, a sixth item (the letter “V”) was added to the similar ensemble that had not been presented in the experiments. Mirroring the experimental procedure, each dissimilar ensemble was used for list construction during half of the simulation replications.

All predictions were based on the average of 1,000 replications at each step during parameter estimation (with the exception of the primacy model, which was run for 10,000 replications because its predictions were found to be more variable than those of the other two models).

All simulations used a maximum likelihood approach (assuming a Gaussian distribution of the to-be-fitted means) to estimate parameters. In order to minimize the total number of parameters to be estimated, we first examined the relationship between the results for the three experiments to determine if any of them could be fit jointly. Particularly, it was noted that Experiment 1 and Experiment 3 used similar methodologies (particularly in the instruction to read letters silently) and gave similar levels of performance; in contrast, the procedure for Experiment

2 required participants to read letters aloud, and gave a higher level of performance than Experiments 1 and 3. Accordingly, parameters were estimated jointly for Experiments 1 and 3, and separately for Experiment 2, using the standard SIMPLEX algorithm which sought to maximize the log likelihood:

$$\ln L = -n/2 \times \ln(SSD/n),$$

where n refers to the number of cell means being fitted, and SSD refers to the sum of squared deviations between those means and the model predictions.

The choice to fit the combined data from Experiments 1 and 3 was confirmed by comparing the combined fit to the goodness-of-fit obtained by fitting Experiments 1 and 3 individually using Akaike's Information Criterion (*AIC*; Akaike, 1974; see also Burnham & Anderson, 2002). The *AIC* corrects the maximized log likelihood for the degrees of freedom of the model, as reflected in the number of free parameters that must be estimated. Formally, the *AIC* is given by:

$$AIC = -2 \ln L + 2K$$

where K is the number of free parameters involved in maximizing $\ln L$. The *AIC* thus trades off goodness-of-fit ($\ln L$) against lack of parsimony (K). In the present case, fitting each study separately requires an extra 4 free parameters compared to the combined fit. A comparison of the joint and individual fits for Experiments 1 and 3 (where $\ln L$ was calculated using only those list types that are common to both experiments; i.e., DDDDDD, SSSDSS, and SDSDSD) showed that for all three models, the joint fits gave a more parsimonious account of the data (i.e., a lower *AIC* value) than the individual fits.

Table 3 summarizes the simulation results and estimated parameter values. We created two numerical measures that captured the essential aspects of the data and predictions; namely, the extent of the mixed-list advantage for dissimilar items and the extent of zig-zagging in the alternating lists, respectively. The first measure was constructed by following relevant precedent

in research on isolation effects (Bone & Goulet, 1968) and consisted of a standardized mixed-vs.-pure index:

$$mvp = [\sum (mixed-pure)/(1-pure)]/n, \quad (1)$$

where *mixed* refers to mean performance for a D item on a mixed list (alternating or isolate) and *pure* to mean performance for the D item in the corresponding serial position on the pure D list, and where the sum is taken across all serial positions and mixed list types and *n* is the number of means being considered. Unlike a net difference measure, this index takes into account baseline performance on the pure D list and thus corrects for primacy in the data. The observed values of *mvp* were .19 for Experiment 1, .16 for Experiment 2, and .34 for Experiment 3. None of our conclusions change qualitatively if a net difference measure is used instead.

The second measure, *ZZ*, consisted of the average difference between each non-terminal dissimilar item and its similar neighbors on the two alternating lists, and is indicative of the extent of zig-zagging of the alternating lists (*ZZ* = .11 for Experiment 1, *ZZ* = .11 for Experiment 2, and *ZZ* = .17 for Experiment 3).

In addition, we present model predictions for Experiment 3 in Figure 4 (because the predictions for Experiments 1 and 2 do not additionally differentiate between the models, graphical representations of those predictions are omitted to conserve space). We first present and discuss the predictions of each of the three models before comparing them in two ways; first, by statistical means and, second, by examining the sensitivity of their predictions to variations in parameter values.

Simulations of the Primacy Model

Implementation

Our implementation of the primacy model followed the description provided by Page and Norris (1998a) and inherited the published values for all fixed parameters. Thus, at encoding the activation of the first list item (P) was set to 11.5, with the activation of each subsequent item decreasing by 1.0. Items were assumed to start decaying exponentially once list presentation (and rehearsal; see Page & Norris, 1998a) was complete, such that at time t the activation of an item, a , was a function of its initial activation a_0 :

$$a = a_0 \times \exp(-D t), \quad (2)$$

where D is a forgetting parameter that was freely estimated.

At each step during retrieval, Gaussian noise with standard deviation N (a free parameter) was added to the activation of all items and the item with the greatest activation was selected as a response candidate and its activation reset to zero to implement response suppression.

The candidate item was forwarded to the response-selection stage where it first entered into a similarity-based confusion process before being compared to a constant omission threshold T ($T = .49$). The similarity-based confusion process was modified from the original formulation of the theory in order to capture the true similarity structure among the stimuli. Unlike the original formulation, in which dissimilar items arbitrarily by-passed this stage, in our implementation any candidate item, whether dissimilar or not, competed for response selection anew with all other list items. The amount of competition offered by each list item was determined by that item's similarity to the candidate weighted by the original primacy gradient. Specifically, given a candidate item i , the activation of each competitor j was computed as:

$$a_j = a_{0j} \times \exp(-S d_{ij}^2), \quad (3)$$

where a_{0j} is the original activation according to the primacy gradient, S is a free parameter, and d_{ij} the Euclidean distance within the MDS solution (see Table 2) between the candidate and that particular competitor. The final activations a_j were perturbed by adding Gaussian noise with standard deviation M (a free parameter), before the maximally active item was chosen as the final response. If that item's activation exceeded T (after addition of Gaussian noise with standard deviation M), it was produced as the model's response (and its activation reset to 0), otherwise an omission was recorded.

Results

When applied to Experiments 1 and 3, the primacy model produced the zig-zagged serial position curves for the alternating lists ($ZZ = .11$ and $.13$, respectively) that were in qualitative agreement with the data. However, the model's predicted mixed-list advantage ($mvp = .10$ and $.09$, respectively) fell short of the observed values ($.19$ and $.34$). Panel A in Figure 4 shows the model's predictions for Experiment 3 and confirms that the mixed-list advantage is qualitatively under-predicted.

For Experiment 2, the model fared worse with a near-zero—and indeed slightly negative—predicted mvp and ZZ . Inspection of the predicted serial position curves (not shown here) revealed that all list types were nearly identical, with performance rather high across all serial positions and only weak primacy.

It is clear that the primacy model cannot produce a noticeable mixed-list advantage and considerably under-predicts performance on the isolate items. The lack of a mixed-list advantage is unsurprising, given that the model was explicitly designed to account for its absence—what is noteworthy instead, is that a very small advantage can be obtained by presenting the model with a more realistic similarity representation than in the original simulations by Page and Norris (1998a). It is also noteworthy that although the primacy model is in principle capable of

producing a convincing zig-zag serial position curve for mixed lists, when fit to the data of Experiment 2, the model was unable to produce any zig-zag at all and instead produced a flat serial position curve for the mixed lists. It follows that the zig-zag for alternating lists does not arise from a deep property of the model but is highly parameter dependent. We explore this parameter-dependency further after we have presented all simulation results.

Simulations of SIMPLE

Simulation Input

For the simulations of SIMPLE, the coordinates in Table 2 were linearly rescaled to set the minimum (across all ensembles and dimensions) equal to 0. Stimuli were represented as points in a four-dimensional space consisting of time plus the three rescaled phonological dimensions.

Implementation

We implemented the three core assumptions of SIMPLE: (1) items were represented by their position within a multi-dimensional psychological space. (2) The similarity between any two items in memory was a reducing function of the distance separating them in that psychological space. (3) The probability of recalling a given item was inversely proportional to that item's summed similarity to all other potentially recallable items.

Encoding in a multi-dimensional space. The remembered location along the temporal dimension was taken to serve as retrieval cue, with the relative importance of the temporal dimension at retrieval governed by the free parameter wt . The weight on each of the phonological dimensions was equal to $(1-wt)/3$. The distance (D) between any two items I and j was thus given by:

$$D_{i,j} = wt|TD_i - TD_j| + (1-wt)/3|P_{1i} - P_{1j}| + (1-wt)/3|P_{2i} - P_{2j}| + (1-wt)/3|P_{3i} - P_{3j}|, \quad (4)$$

where TD_i is the (logarithmically transformed) temporal distance of stimulus i from the time of retrieval, and P_{ki} is the phonological coordinate value on dimension k for item i . Although it would be in principle possible to use a separate dimensional weight for each phonological dimension, thus stretching (or shrinking) the psychological space according to the relevance of each dimension, the requirement to equalize the number of free parameters between models ruled out this option. Similarly, although recent applications of SIMPLE have included a positional dimension in addition to time (e.g., Lewandowsky, Duncan, & Brown, 2004; Lewandowsky, Brown, Wright, & Nimmo, 2006), we did not include a positional dimension here because it does not alter the predictions of SIMPLE much unless the temporal parameters of encoding or retrieval are explicitly manipulated.

Similarity-distance metric. The similarity of any two items in memory was a reducing exponential function of the distance between them in psychological space. Thus:

$$\eta_{i,j} = e^{-cD}, \quad (5)$$

where $\eta_{i,j}$ is the similarity between items i and j and D is the distance between them. The free parameter c governs the rate at which similarity falls off with distance, with the steepness of the similarity gradient increasing with the magnitude of c . In consequence, items in close psychological proximity have a similarity approaching unity, whereas items that are more psychologically distant have a similarity that, in the extreme, tends towards zero.

Similarity determines recall. The probability of recalling a given item $P(R_i)$ was inversely proportional to the summed similarity of that item to every other potentially recallable response:

$$P(R_i) = \frac{1}{\sum_{k=1}^n (\eta_{i,k})}, \quad (6)$$

where n is the number of items that could be recalled (here, the number of list items).

Omissions were assumed to arise from thresholding of low retrieval probabilities (as calculated by the preceding equation). Brown et al. (2002) implement this threshold mechanism using a sigmoid function such that if P was the recall probability as just defined, it was transformed to become P^* as follows:

$$P^* = \frac{1}{1 + e^{-s(P-t)}}, \quad (7)$$

where free parameters t and s refer to the absolute value and to the noisiness of the threshold, respectively (see Brown et al., 2002, for a complete account).

Results

For Experiments 1 and 3, the theory predicted a small mixed-list advantage ($mvp = .10$ and $.11$, respectively) which, like the predictions of the primacy model, fell considerably short of the observed values ($.19$ and $.34$). Unlike the primacy model, SIMPLE failed to predict any noticeable zig-zagging of the alternating lists ($ZZ = .002$ and $.02$), which put the theory qualitatively at odds with the data. Panel B in Figure 4 shows the model's predictions for Experiment 3.

For Experiment 2, by contrast, the predicted ZZ was considerably larger ($.13$) and in line with the data ($.11$). However, the more plausible value of ZZ was not accompanied by a sizable mvp , which again fell short of the observed value ($.09$ vs. $.16$). Moreover, predicted performance on the pure similar list, which could not contribute to the fit because it was absent in the experiments, was unreasonably low. Unlike the predictions for Experiment 3, performance on the pure similar list was predicted to be some 40% below that of the pure dissimilar list for the last three serial positions: This predicted difference exceeds that commonly obtained in the data (e.g., Experiment 2 of Farrell & Lewandowsky, 2003).

Discussion

To examine the reasons for SIMPLE's inability to produce a consistent zig-zag for alternating lists, we explored the behavior of the model using a readily tractable single similarity dimension, based on a one-dimensional MDS solution for the confusion matrix of Hull (1973). This exploration revealed that the extent of zig-zag in the serial position curves for alternating lists was dependent on the generalization parameter, c .

When c was moderately large ($c \gg 10$), no zig-zag of any kind was present and S items on alternating lists were recalled with the same accuracy as D items. This reflects the fact that with large values of c (i.e., very steep generalization gradients), recall of any given item is affected only by its immediate neighbors whereas more remote items matter very little. Hence, S and D items on alternating lists are recalled equally well, because any S item is as distinct from its neighboring D items as each of those D items is from its S neighbors.

This symmetry is broken if the value of c is lowered ($c \leq 10$) and generalization gradients among items begin to overlap, thus allowing more remote neighbors to contribute to an item's distinctiveness. In that case, S items are at a disadvantage even on alternating lists, because they are closer along the phonological dimension to other S items than the D items are to other dissimilar list members. However, when c is sufficiently low to produce striking zig-zagging on alternating lists, performance on pure S lists suffered even more and rapidly became unreasonably low.

Simulations of SOB

Simulation Input

For SIMPLE, the MDS solution directly translated into a suitable stimulus representation for the simulations. For SOB by contrast, the translation between MDS coordinates and stimulus

representations required several design decisions. We chose to represent items by 150-element vectors that were divided into three segments of 50 elements, each segment corresponding to a dimension in the MDS solution.

Item vectors were generated by probabilistic interpolation between two random “anchor” vectors, constructed from the set $\{-1,+1\}$ anew for each replication. Those two anchor vectors represented, respectively, the origin and the maximum possible value of all coordinates. MDS coordinates were rescaled to a value p within the range $0 - 1$, and for any given stimulus, each feature was then sampled from the maximum-coordinate vector with probability p and from the origin with probability $1-p$. This mechanism ensured that, as in SIMPLE, all three dimensions were given equal weight and it ensured that MDS coordinates determined the similarity relationship among stimulus vectors.

Implementation

The present implementation of SOB was an extension of the model presented by Farrell and Lewandowsky (2002) which was first reported by Farrell (2006). The original SOB rested on three principal architectural commitments: (1) Similarity-sensitive encoding as described previously; (2) implementation of response suppression by the same process that governed encoding; and (3) dynamic iterative deblurring of partial images retrieved from memory. Those three principles enabled the model, respectively, to compute a primacy gradient (and hence primacy in the serial position curve) endogenously, to produce recency endogenously, and to provide a process description of response selection. Here, we retained the first two commitments but relaxed the third one for reasons of computational convenience and efficiency (the present simulations could not have been conducted within a reasonable time on available equipment if dynamic iterative deblurring had been implemented).

In addition, in the current version of SOB, items could be similar to each other. The introduction of similarity between items is a non-trivial extension of the original model which, in turn, necessitated the use of a positional marker to represent order.

Architecture. The network consisted of two layers; an input layer ($N = 16$) used to represent positional markers, and an output layer ($N = 150$) representing list items, where N refers to the number of units in each layer. The two layers were fully interconnected by a weight matrix \mathbf{W} .

Positional markers. This generalized version of SOB incorporated a positional marker for each list position, as is commonly assumed in models of memory (e.g., Brown et al., 2000; Burgess & Hitch, 1999). The similarity between any two positional markers was an exponential function of their absolute separation in list positions; that is,

$$\cos(\mathbf{p}_i, \mathbf{p}_j) = t_c^{|i-j|}, \quad 8$$

where i and j are the positions of the i th and j th items, \mathbf{p}_i and \mathbf{p}_j are distributed vectors representing positional markers at those positions, and t_c was set to a constant .25. Markers were generated from a weighted combination of orthogonal Walsh vectors of dimensionality 16, with the weights set to satisfy the similarity constraint embodied in Equation 8. In consequence, only positional information was represented and the temporal separation between items at encoding was ignored, as is suggested by recent research on temporal isolation effects (e.g., Lewandowsky et al., 2006; Nimmo & Lewandowsky, 2005).

Encoding. At list presentation, incoming items were associated with successive positional markers using standard Hebbian learning (see, e.g., Anderson, 1995):

$$\Delta \mathbf{W}_i = \eta_e(i) \mathbf{v}_i \mathbf{p}_i^T, \quad (9)$$

where \mathbf{W} is the matrix of positional marker-item weights, \mathbf{v} is the vector representing the i th presented item, and \mathbf{p} is a positional marker for the i th serial position. The learning rate η_e for the

i th association, $\eta_e(i)$, was calculated using the *energy* between the association to be learned and the associations stored in the weight matrix up to that point:

$$\eta_e(i) = \begin{cases} 1, & i = 1 \\ -\phi_e / E_i, & i > 1 \end{cases} \quad (10)$$

where ϕ_e is a free parameter, and E_i , the energy of the i th association, is given by

$$E_i = -\mathbf{v}_i^T \mathbf{W}_{i-1} \mathbf{p}_i. \quad (11)$$

(see Farrell & Lewandowsky, 2002). The use of energy to compute the weighting of incoming information implements a core principle of SOB responsible for the similarity-sensitive encoding explored here.

Energy is a measure of the consistency between a new to-be-learned association and the earlier associations that have already been learned on that trial. Another way of looking at energy here is as a measure of the novelty of an item in the context of a particular positional marker. This relationship quantitatively encapsulates the idea that encoding of similar items will result in more redundant storage, and that scaling the learning rate by the novelty of a position-item pairing will limit such redundancy; more importantly, as discussed earlier, this learning scheme drives the enhancing effects of similarity in the model.

Retrieval. Retrieval in the network consists of stepping through the positions and using the positional markers to cue for their associated items. For position i , retrieval was effected by placing positional marker \mathbf{p}_i across the input layer, and updating the item unit activations:

$$\mathbf{v}_i' = \mathbf{W}_i \mathbf{p}_i, \quad (12)$$

where the vector produced, \mathbf{v}_i' , was generally a “noisy” version of \mathbf{v}_i , containing a blend of \mathbf{v}_i and the other items on the list, according to the overlap between the positional markers for those other items, and the positional marker \mathbf{p}_i . The probability of recalling an item \mathbf{v}_j was then

determined by assuming that the noisy vector was matched to a pool of possible responses, in this case the experimental vocabulary. Formally, the probability of recalling an item \mathbf{v}_v from the experimental vocabulary was

$$P(\mathbf{v}_v) = \frac{s(\mathbf{v}_i', \mathbf{v}_v)}{\sum_{k=1}^n s(\mathbf{v}_i', \mathbf{v}_k)}, \quad (13)$$

where n was the number of items in the experimental vocabulary and the function s , the similarity between the noisy vector and an item vector, was calculated as

$$s(\mathbf{v}_i', \mathbf{v}_k) = \exp[-cD(\mathbf{v}_i', \mathbf{v}_k)^2]. \quad (14)$$

The Euclidian distance measure D was weighted by the free parameter c , and was normalized by subtracting the minimum distance, across all the n items, from the distance for each item; this ensured that the obtained distances were close to 0, as otherwise, due to the amplification inherent in Equation 12, the distances were all extremely large and did not allow computation of exponential scaling in Equation 14. An item was then selected for recall by randomly sampling an item according to Equation 13.

Once an item had been recalled, it was suppressed by adjusting the weights between the positional marker layer and the item layer, according to

$$\Delta \mathbf{W}_j = \eta_s(j) \mathbf{v}_{o,j} \mathbf{p}_j^T, \quad (15)$$

where j is the output position, and $\mathbf{v}_{o,j}$ is the item recalled at the j th position; Equation 15 is identical to the Hebbian learning used at encoding, except that the learning rate was negative, thus causing “anti-learning” (e.g., Anderson, 1991) of the weights and attenuation of item representations. To ensure that the extent of response suppression approximately matched that of

learning, the learning rate for suppression, η_s , was also determined from the energy of the recalled item with respect to the association matrix \mathbf{W} and the positional marker \mathbf{p}

$$E_j = -\mathbf{v}_{o,j}^T \mathbf{W}_{j-1} \mathbf{p}_j. \quad (16)$$

The response suppression rate was given by

$$\eta_s(j) = \frac{-E_j}{\phi_s E_1}, \quad (17)$$

where ϕ_s was a free parameter, and E_1 was the energy of the item recalled at the first output position (see Farrell & Lewandowsky, 2002); this term reduced to $1/\phi_s$ for the first output position (where $j=1$).

Reflecting a fairly general consensus in the short-term memory literature (e.g., Brown et al., 2000; Cowan, Saults, Elliott, & Moreno, 2002; FitzGerald & Broadbent, 1985; Oberauer, 2003), we also assumed that retrieval of an item was accompanied by output interference. Output interference was modeled by adding Gaussian noise with standard deviation N_o (a free parameter) to each weight in \mathbf{W} after each retrieval.

Results

When applied to Experiments 1 and 3, SOB slightly over-predicted the extent of the mixed-list advantage for Experiment 1 ($mvp = .26$) but mirrored the observed value closely for Experiment 3 ($mvp = .36$). Panel C in Figure 4 shows the predicted serial position curves for Experiment 3. Comparison with the data in the right-hand panel of Figure 3 shows that SOB handled the data of Experiment 3 better than the other models.

The predicted mvp for Experiment 2 also fell into the appropriate range ($mvp = .20$). Across both fits, the model predicted an appropriate amount of zig-zag for all three experiments.

Discussion

The behaviour of SOB, which mirrors the data well, reflects SOB's assumption that encoding is sensitive to the similarity of list items. In particular, SOB predicts the enhancing effects of similarity for dissimilar items due to the decreased encoding given to the similar items. Thus, SOB does not predict that dissimilar items receive any more encoding on mixed lists than they do on pure dissimilar lists (contrasting SOB with other novelty-based explanations for mixed-list effects: see, e.g., Fabiani & Donchin, 1995); instead, the competing similar items on mixed lists receive less encoding than dissimilar items on pure lists, thus giving dissimilar items on mixed lists a *relative* advantage at recall.

A Commonality among Models: Similar Items on Mixed-Lists

Before we turn to a statistical comparison of the models' accounts, it is worth considering one other aspect of the data; viz. the effect of mixed-lists on the recall of similar items. Inspection of Figures 2 and 3, along with previous mixed-list results (e.g., Farrell & Lewandowsky, 2003; Henson et al., 1996) shows that similar items are recalled better if they are placed on alternating lists rather than on isolate lists or pure S lists. Notably, SIMPLE, the primacy model, and SOB all predict this effect (see Figure 4). This prediction arises from the standard property of all three models that—irrespective of their assumptions about encoding processing—similar items are more confusable at retrieval than dissimilar items.

The MDS solution underlying the modeling facilitates an intuitive understanding of the consequences of this retrieval assumption: When similar items are placed on mixed lists, as opposed to pure lists, their average pairwise similarity to all other list items will necessarily be lower. In SIMPLE, this reduced pairwise similarity leads to a reduction of the denominator of the choice rule involved in the selection of items (see Equation 6). In SOB, lower pairwise similarity likewise facilitates item choice for the same reason (see Equation 13), although SOB

additionally involves the energy-gated encoding processes that were described earlier. In the primacy model, the reduced pairwise similarity lowers the average activation of competing items (see Equation 3). The shared retrieval assumption thus enables all three models to handle some aspects of our data; we next show that the models can nonetheless be statistically differentiated.

Statistical Model Comparison

Visual comparison of the predictions in Figure 4 to the data in Figure 3 suggests that SOB captured the main aspects of the data better than the other two models, an impression confirmed by the descriptive summary statistics (*mvp* and *ZZ*) reported in Table 3. We next examine the differences between model fits at a quantitative level, using the log likelihood statistics (*lnL*) that were obtained during parameter estimation and that summarize the goodness of fit of each model.

Table 4 summarizes the log likelihood statistics and their relationship between the three models. Note that because the number of free parameters is equal for all models, comparison can rely on the raw log likelihoods without correcting for the various models' degrees of freedom (e.g., by AIC or BIS; cf. Wagenmakers & Farrell, 2004).

The column labeled ΔlnL contains the pairwise differences in the log likelihood statistic (*lnL*) between each given model and its best-fitting counterpart (this difference is necessarily zero for the best-fitting model; viz. SOB). Because those differences are taken between logarithms of likelihoods, entries in this column correspond to (the logarithm of) the likelihood ratios between the competing models. The final column, then, reports the actual likelihood ratios, denoted λ and obtained by exponentiation of the previous column. Likelihood ratios are a statistic that reflects the relative likelihood of the observed data, given two competing models (see, e.g., Glover & Dixon, 2004; Royall, 1997).

The final column reveals that for both fits, the data were considerably less likely to have been generated by the primacy model or SIMPLE than by SOB. Clearly, the magnitudes of the

likelihood ratios unequivocally suggest that SOB is the quantitatively preferred model. What remains to be seen is the extent to which this preference is tied to the specific parameters estimated here or to the models' architecture.

Parameter Sensitivity Analysis

We next examined the sensitivity of the models' predictions to variations in the best-fitting parameter values (cf. Li, Lewandowsky, & DeBruener, 1996). Parameter sensitivity analysis can ascertain whether the predictions of a model derive from its deep architectural principles or are tied to specific parameter values. For example, a model might successfully fit the data overall but might nonetheless fail to capture important aspects of the results if its parameters are slightly perturbed; conversely, a model that fits the data poorly may nonetheless capture qualitative aspects of the results if goodness-of-fit is relaxed and the parameters are altered slightly.

We investigated these issues by performing a grid-search through the region of the parameter space circumscribed by the best-fitting points associated with each experiment. Because emphasis here was on a model's flexibility and sensitivity to parameter values rather than its quantitative fit, we no longer combined experiments but obtained best-fitting parameter estimates (not reported in detail) for each model and each experiment separately. This enabled each model to accommodate at least part of the data with the smallest permissible number of constraints. Having thus obtained three points in parameter space, we created a grid by dividing the range between each parameter's maximum and minimum estimate across the three experiments into 5 steps, which were then factorially combined across the four parameters to yield a total of 625 cells. This grid of 625 cells spanned the region of the parameter space traversed by each model when given maximal flexibility to fit the various data sets.

We obtained predictions from each model (based on the usual number of replications) for each of the 625 cells of the grid. The distributions of predicted *mvp* and *ZZ* values are shown in

Figures 5 and 6, respectively. In each figure, the thick vertical lines represent the values obtained in the three experiments under consideration. It is clear from the figures that the distributions for SOB comfortably straddle at least two of the three observed values (narrowly missing one *mvp* and one *ZZ* value), whereas the centers of gravity of the distributions for the primacy model and SIMPLE clearly fall outside the range of observed values. This suggests that within the section of parameter space identified by the present experiments, SOB *necessarily* predicts a mixed-list advantage accompanied by zig-zagging of the alternating lists, whereas the other two models consistently and nearly inevitably under-predict both features of the data.

It is important to note that even though the tails of the distributions of SIMPLE and the primacy model cover the empirical values, this does not imply that the models can provide a quantitative account of the results: Goodness-of-fit was irrelevant in the parameter-sensitivity analysis and hence, while a model's *mvp* or *ZZ* might match an empirical value, it does not follow that the associated goodness-of-fit would be at all acceptable. Indeed, the earlier quantitative comparison of the three models showed that SIMPLE and the primacy model provide an inferior account of the data. The outcome of the sensitivity analysis confirms that this earlier conclusion was not based on idiosyncratic parameter estimates but represented deep properties of each model's architecture.

Finally, the modeling results also allay possible concerns about the magnitude of the mixed-list advantage in the present experiments. The fact that the predictions of two of the models fell systematically short of the observed values clarifies that the effect, though arguably small in magnitude when considered on its own, was sufficiently large to have the desired theoretical impact.

CONCLUSIONS

We extended the generality of the mixed-list advantage for dissimilar items in short-term serial recall. Unlike related previous experiments (Farrell & Lewandowsky, 2003; Henson et al., 1996), in the present studies we randomized all mixed list types to minimize any effects due to potential encoding strategies that are beyond the purview of serial recall models. In addition, across the two experiments we reversed the balance between the total number of similar and dissimilar items; whereas there were more similar than dissimilar tokens overall in Experiment 1, the reverse was true for Experiment 2. Notwithstanding those departures from previous methodologies, we observed a mixed-list advantage for dissimilar items in both experiments.

It is nonetheless possible to argue that the mixed-list advantage might arise from some strategic effect, such as grouping of the alternating lists into a set of similar items and a set of dissimilar items. However, this strategy explanation runs afoul of the fact that no mixed-list advantage was observed with alternating lists by Henson et al. (1996), despite the fact that their presentation regime was identical to our Experiment 2. A strategy account would also struggle to accommodate the converse pattern; namely, the consistent presence of a mixed-list advantage in the present experiment and the studies of Farrell (2006) and Farrell and Lewandowsky (2003) despite a variety of methodological differences that would be expected to affect people's strategies. We therefore suggest that any alternative account that appeals to people's strategies must await further specification before it can present a viable alternative to our conclusion that similarity among list items determines their encoding into memory.

Our results challenge dual-stage models of serial recall, some of which were specifically designed to accommodate the absence of a mixed-list advantage (e.g., Page & Norris, 1998a). Two-stage models postulate that the mechanism responsible for ordering items in a sequence must be independent of the phonological similarity of those items. The current findings, along

with those of Farrell and Lewandowsky (2003) and Farrell (2006), remove a key piece of support for dual-stage models and, by implication, suggest that a single-stage approach might be preferable.

This conclusion was supported by our simulations which showed that neither the primacy model, in which ordering is achieved in a stage upstream from phonological confusions, nor SIMPLE, in which the temporal dimension is independent of the dimensions representing phonological similarity, could satisfactorily account for the observed mixed-list effects. The crucial architectural similarity—that is, the presence of two independent stages—between the primacy model and Henson’s (1998) SEM and the Burgess and Hitch (1999) model implies that the latter two models are also challenged by our results. This was confirmed by Farrell (2006), who showed that SEM was unable to produce a mixed-list advantage for dissimilar items in related circumstances. Although it would be premature to dismiss two-stage models on the basis of a single empirical effect, we argue that the mixed-list advantage provides an empirical constraint that future revisions of two-stage theories must strive to accommodate.

We find it more parsimonious, however, to conclude from the present modeling that phonological similarity must exert an effect at the encoding stage and that, by implication, single-stage models such as SOB represent a preferable line of theorizing. The conclusion that similarity effects are located at an encoding stage is not entirely new: Chechile (1977) came to a similar conclusion on the basis of the multinomial modeling of an experiment in which recall and recognition tests of similar or dissimilar trigrams were randomly intermixed. In Chechile’s study, 95% of the variance in performance (relative to a baseline) was accounted for by storage processes, with confusions at retrieval being responsible for the remaining small fraction of variance. This early finding was echoed very strongly in the present results, and the consistent

mixed-list advantage for dissimilar items therefore places clear constraints on the future development of models of serial recall from short-term memory.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.
- Anderson, J.A. (1991). Why, having so many neurons, do we have so few thoughts? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honour of Bennet B. Murdock* (pp. 477-507). Hillsdale, NJ: Erlbaum.
- Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.
- 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*, *20A*, 249-263.
- Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, *13*, 80-97.
- Bone, R. N., & Goulet, L. R. (1968). Serial position and the von Restorff isolation effect. *Journal of Experimental Psychology*, *76*, 494-496.
- Bridges, A. M., & Jones, D. M. (1996). Word dose in the disruption of serial recall by irrelevant speech: Phonological confusions or changing-state? *Quarterly Journal of Experimental Psychology*, *49A*, 919-939.
- Brown, G. D. A., & Chater, N. (2001). The chronological organization of memory: Common psychological foundations for remembering and timing. In C. Hoerl & T. McCormack (Eds.), *Time and memory: Issues in philosophy and psychology* (pp. 77-110). Oxford, UK: Oxford University Press.

- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127-181.
- Brown, G. D. A., Neath, I., & Chater, N. (2002). *A local distinctiveness model of scale-invariant memory and perceptual identification*. Manuscript submitted for publication.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551-581.
- Burnham, K.P. & Anderson, D.R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer-Verlag.
- Chechile, R. (1977). Storage-retrieval analysis of acoustic similarity. *Memory & Cognition, 5*, 535-540.
- Cimbalo, R. S., Nowak, B. I., & Soderstrom, J. A. (1981). The isolation effect in children's short-term memory. *Journal of General Psychology, 105*, 215-223.
- Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel word lists. *Memory & Cognition, 21*, 539-545.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75-84.
- Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*, 153-177.
- Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 224-240.
- Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language, 55*, 587-600.

- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*, 59-79.
- Farrell, S., & Lewandowsky, S. (2003). Dissimilar items benefit from phonological similarity in serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 838-849.
- FitzGerald, P., & Broadbent, D. E. (1985). Order of report and the structure of temporary memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 217-228.
- Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review*, *11*, 791-806.
- Henry, L. A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *Quarterly Journal of Experimental Psychology*, *43A*, 35-52.
- Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End model. *Cognitive Psychology*, *36*, 73-137.
- 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.
- Hull, A. J. (1973). A letter-digit matrix of auditory confusions. *British Journal of Psychology*, *64*, 579-585.
- Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 586-594.

- Kelley, M. R., & Nairne, J. S. (2001). Von Restorff revisited: Isolation, generation, and memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 54-66.
- Kruskal, J. B. (1964). Nonmetric Multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115-129.
- Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language*, *54*, 20-38.
- Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review*, *11*, 771-790.
- Lewandowsky, S. & Murdock, B.B. (1989). Memory for serial order. *Psychological Review*, *96*, 25-57.
- Li, S.-C., Lewandowsky, S., & DeBrunner, V. E. (1996). Using parameter sensitivity and interdependence to predict model scope and falsifiability. *Journal of Experimental Psychology: General*, *125*, 360-369.
- Lippman, L. G. (1980). Perceptual isolation and position judgement. *American Journal of Psychology*, *93*, 683-695.
- Murdock, B. B., & Kahana, M. J. (1993a). An analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 689-697.
- Murdock, B. B., & Kahana, M. J. (1993b). List-strength and list-length effects: Reply to Shiffrin, Ratcliff, Murnane, and Nobel (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1450-1453.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251-269.

- Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition*, 27, 45-53.
- Neath, I. (1999). Modelling the disruptive effects of irrelevant speech on order information. *International Journal of Psychology*, 34, 410-418.
- Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, 7, 403-423.
- Nimmo, L. M., & Lewandowsky, S. (2005). From brief gaps to very long pauses: Temporal isolation does not benefit serial recall. *Psychonomic Bulletin & Review*, 12, 999-1004.
- Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, 49, 469-483.
- Page, M. P. A., & Norris, D. (1998a). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761-781.
- Page, M. P. A., & Norris, D. (1998b). 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). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Royall, R.M. (1997). *Statistical evidence: A likelihood paradigm*. London: Chapman & Hall.
- Shiffrin, R., Ratcliff, R., Murnane, K., & Nobel, P. (1993). TODAM and the list-strength and list-length effects: Comment on Murdock and Kahana (1993a). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1445-1449.
- Solso, R. L., & Juel, C. L. (1980). Positional frequency and versatility of bigrams for two-through nine-letter English words. *Behavior Research Methods & Instrumentation*, 12, 297-343.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights.

Psychonomic Bulletin & Review, *11*, 192-196.

Wickelgren, W. A. (1965a). Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *4*, 53-61.

Wickelgren, W. A. (1965b). Short-term memory for phonemically similar lists. *American Journal of Psychology*, *78*, 576-574.

Author Note

Preparation of this paper was facilitated by a Large Grant, a Discovery Grant, and a Linkage International Grant from the Australian Research Council to the first author. During writing, the second author was partly supported by NIMH grant HD MH44640 and NIA grant AG17083-01.

Address correspondence to the first author at School of Psychology, University of Western Australia, Crawley, W.A. 6009, Australia. Electronic mail may be sent to lewan@psy.uwa.edu.au. Personal home page at <http://www.psy.uwa.edu.au/user/lewan/>.

We gratefully acknowledge assistance from Leo Roberts at all stages of manuscript preparation.

Footnotes

¹ An alternative approach to controlling guessing would have involved the use of a large pool of words rather than letters. We opted against this alternative because the modeling reported in this article relies on knowledge of the exact similarity structure among stimuli. Much is known about the psychological representation of phonological similarity among letters (via MDS analysis) whereas comparatively little is known about words.

² The results for the final trial of one participant were missing owing to a copying error. Data for that participant are therefore based on 23 instead of 24 DDDDDD trials.

³ In another analysis, responses to all mixed-list dissimilar items were combined into a composite serial position curve by averaging across alternating and isolate lists at serial positions 4 and 6. The 2×6 (List Type: mixed vs. pure \times Serial Position) within-participants ANOVA revealed main effects of serial position, $F(5, 150) = 39.29$, $MSe = .026$, $p < .0001$, and list type, $F(1, 30) = 7.83$, $MSe = .017$, $p < .01$, but no interaction between the two variables, $F(5, 150) = 1.14$.

⁴ The mixed-list advantage was absent for serial position 5. Inspection of the observations at that serial position showed that for the alternating lists, the distribution across participants was unusually skewed with the median (.75) exceeding the mean (.68) and 4 participants responding with less than .2 accuracy. No other serial position for either list type showed a similar skew.

Table 1. *Average Stimulus Predictability for All Experimental List Types*

| List Type | Dissimilar Ensemble | |
|-------------------------------|---------------------|---------------|
| | H, K, M, Q, R | X, L, Z, W, Y |
| Mixed (e.g., SDSDSD) | 1.88 | 1.74 |
| Single isolate (e.g., SSSDSS) | 1.89 | 1.81 |
| Pure dissimilar (DDDDDD) | 1.50 | |

Table 2. *Three-Dimensional MDS Coordinates for All Stimuli in the Three Ensembles Used in the Experiments.*

| | Dimension | | |
|-----------------------|-----------|-------|-------|
| | 1 | 2 | 3 |
| Similar Ensemble | | | |
| B | 1.21 | 0.32 | -0.59 |
| D | 1.47 | 0.15 | -0.45 |
| G | 1.51 | -0.28 | 0.27 |
| P | 1.08 | 0.60 | -0.29 |
| T | 1.45 | 0.59 | 0.17 |
| V | 1.26 | 0.15 | -0.95 |
| Dissimilar Ensemble 1 | | | |
| H | -1.73 | -0.01 | -0.11 |
| K | -0.70 | 1.35 | 1.06 |
| M | -0.67 | 1.33 | -0.98 |
| Q | 1.06 | -0.78 | 1.08 |
| R | -0.41 | -2.05 | -0.16 |
| Dissimilar Ensemble 2 | | | |
| L | -0.98 | -0.25 | 1.54 |
| W | 0.22 | -1.37 | -1.64 |
| X | -1.49 | -0.20 | -0.86 |
| Y | -0.47 | -1.90 | 0.73 |
| Z | -1.19 | 0.73 | 0.84 |

Table 3. Summary Of Parameter Values And Results For All Simulations.

| Model and Parameters | Experiment 1 (<i>mvp</i> =.19; <i>ZZ</i> =.11) | | | | Experiment 3 (<i>mvp</i> =.34; <i>ZZ</i> =.17) | | | | Experiment 2 (<i>mvp</i> =.16; <i>ZZ</i> =.11) | | | |
|----------------------|--|-----------------------|----------------------|----------------------|--|-----------------------|----------------------|----------------------|--|-----------------------|----------------------|----------------------|
| | <i>M</i> | <i>D</i> | <i>N</i> | <i>S</i> | <i>M</i> | <i>D</i> | <i>N</i> | <i>S</i> | <i>M</i> | <i>D</i> | <i>N</i> | <i>S</i> |
| Primacy Model | | | | | | | | | | | | |
| | 1.03 | .37 | .14 | .81 | | | | | 1.39 | .07 | .43 | 1.99 |
| <i>mvp</i> | .10 | | | | .09 | | | | | -.002 | | |
| <i>ZZ</i> | .11 | | | | .13 | | | | | -.005 | | |
| SIMPLE | | | | | | | | | | | | |
| | <i>c</i> | <i>wt_t</i> | <i>t_o</i> | <i>s_o</i> | <i>c</i> | <i>wt_t</i> | <i>t_o</i> | <i>s_o</i> | <i>c</i> | <i>wt_t</i> | <i>t_o</i> | <i>s_o</i> |
| | 20.07 | .96 | .89 | 19.74 | 10.76 | .32 | .86 | 13.46 | | | | |
| <i>mvp</i> | .10 | | | | .11 | | | | | .09 | | |
| <i>ZZ</i> | .002 | | | | .02 | | | | | .13 | | |
| SOB | | | | | | | | | | | | |
| | ϕ_e | ϕ_s | <i>N_o</i> | <i>c</i> | ϕ_e | ϕ_s | <i>N_o</i> | <i>c</i> | ϕ_e | ϕ_s | <i>N_o</i> | <i>c</i> |
| | 207.36 | .45 | 1.64 | .30 | 209.15 | .85 | .80 | .26 | | | | |
| <i>mvp</i> | .26 | | | | .36 | | | | | .20 | | |
| <i>ZZ</i> | .16 | | | | .14 | | | | | .10 | | |

Table 4. *Comparison of Goodness of Fit For All Models.*

| Experiment and Model | $\ln L$ | $\Delta \ln L$ | λ |
|----------------------|---------|----------------|----------------------|
| Experiments 1 & 3 | | | |
| Primacy model | 164.53 | 5.9 | 365.04 |
| SIMPLE | 151.42 | 19.01 | 1.8×10^8 |
| SOB | 170.43 | 0 | |
| Experiment 2 | | | |
| Primacy model | 53.66 | 23.72 | 2.0×10^{10} |
| SIMPLE | 60.41 | 16.97 | 2.3×10^7 |
| SOB | 77.38 | 0 | |

Figure Captions

- Figure 1. Summary of transposition rates for isolate items in all experiments reported by Farrell and Lewandowsky (2003). Performance is shown for items on mixed lists (mixed D) and for items in corresponding serial positions on pure lists (pure D). Mixed lists include alternating (SDSDSD) and single-isolate lists (e.g. SDSSSS). APA Copyright permission to be obtained.
- Figure 2. Serial position curves for all list types in Experiment 1 (left-hand panel) and composite serial position curves for dissimilar items taken from mixed and pure lists for the same experiment (right-hand panel).
- Figure 3. Serial position curves for all list types in Experiment 2 (left-hand panel); composite serial position curves for dissimilar items taken from mixed and pure lists for the same experiment (center panel); and serial position curves for all list types in Experiment 3 of Farrell and Lewandowsky (2003).
- Figure 4. Predicted serial position curves for all models for all list types in Experiment 3 of Farrell and Lewandowsky (2003) plus pure similar lists. All predictions are based on best-fitting parameter estimates obtained by maximizing log likelihood for Experiments 1 and 3 jointly. Panel A shows the predictions of the primacy model; Panel B shows the predictions of SIMPLE; Panel C the predictions of SOB.
- Figure 5. Distribution of predicted values for *mvp* from a parameter-sensitivity analysis for all models (see text for details). From top to bottom, the predictions are for the primacy model, SIMPLE, and SOB. The values observed in the present Experiments 1 and 2, and Experiment 3 of Farrell and Lewandowsky (2003), are represented by the correspondingly numbered thick vertical lines.

Figure 6. Distribution of predicted values for ZZ from a parameter-sensitivity analysis for all models (see text for details). From top to bottom, the predictions are for the primacy model, SIMPLE, and SOB. The values observed in the present Experiments 1 and 2, and Experiment 3 of Farrell and Lewandowsky (2003), are represented by the correspondingly numbered thick vertical lines.

Figure 1

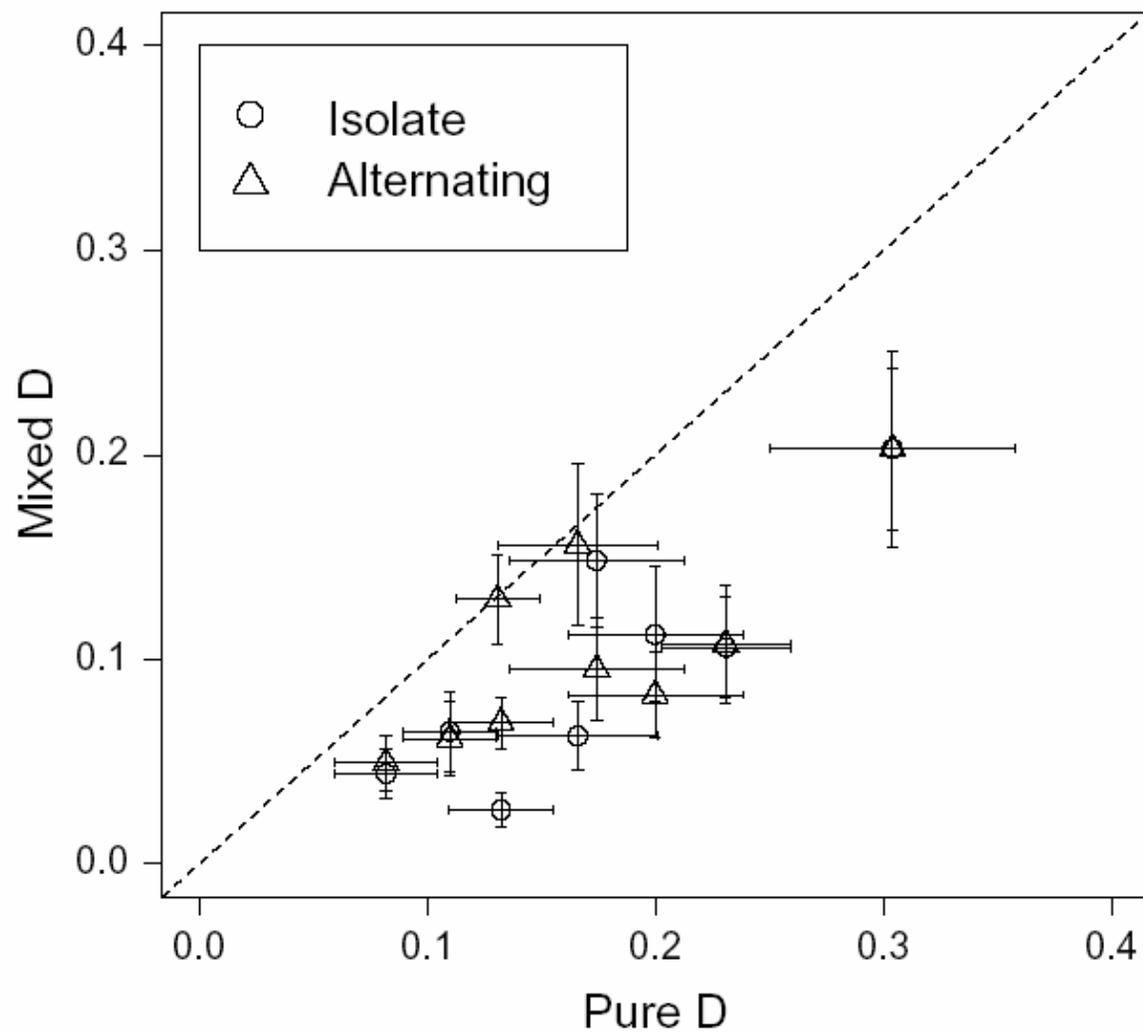


Figure 2

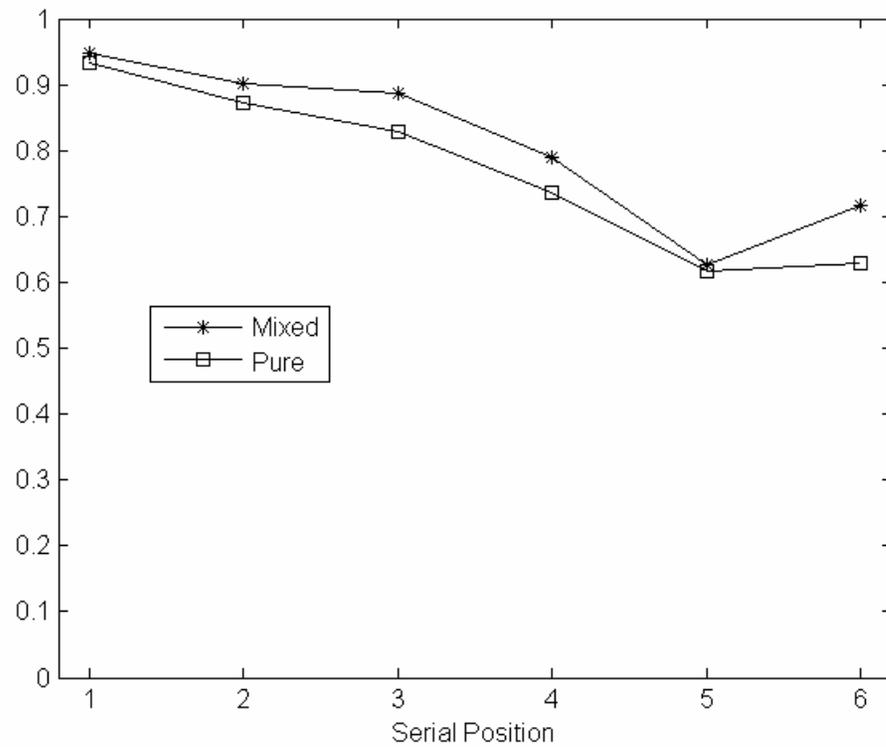
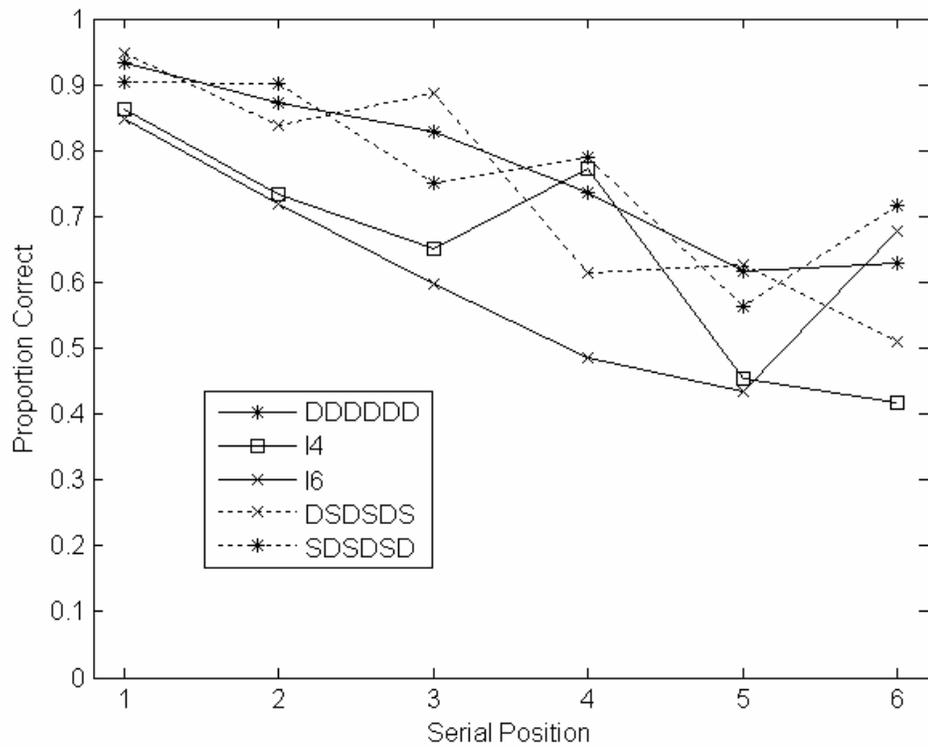


Figure 3

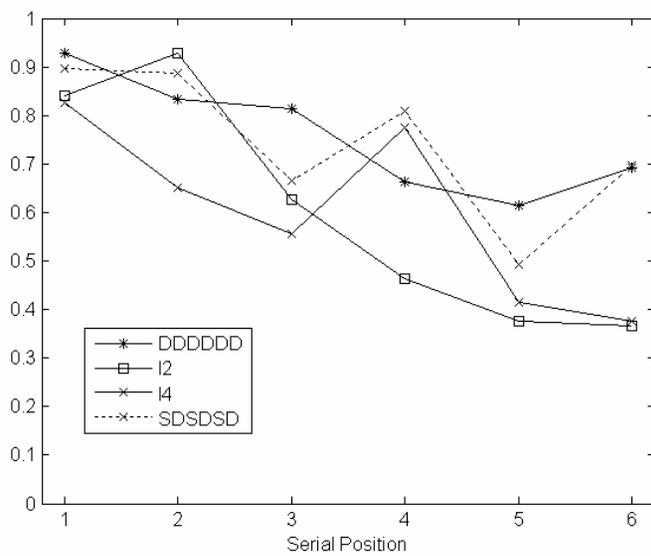
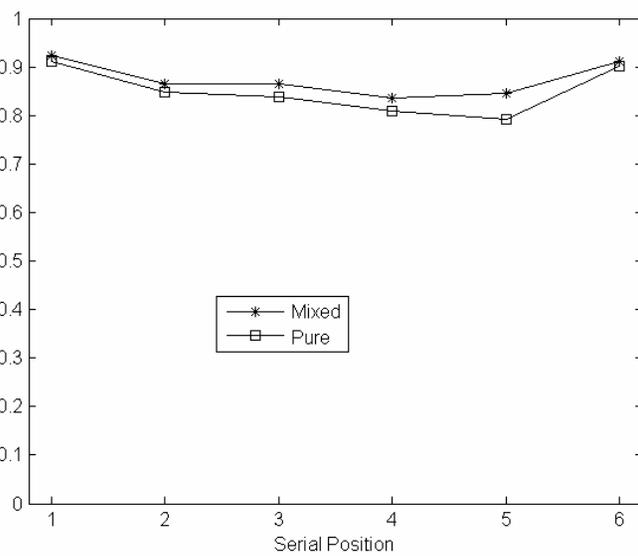
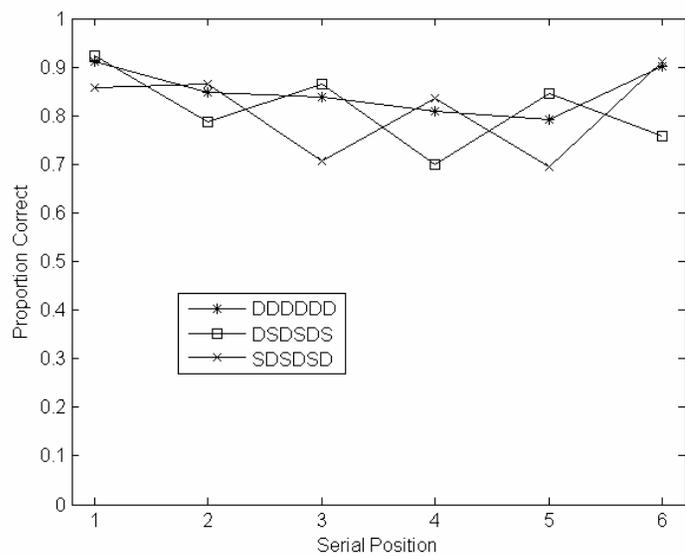


Figure 4

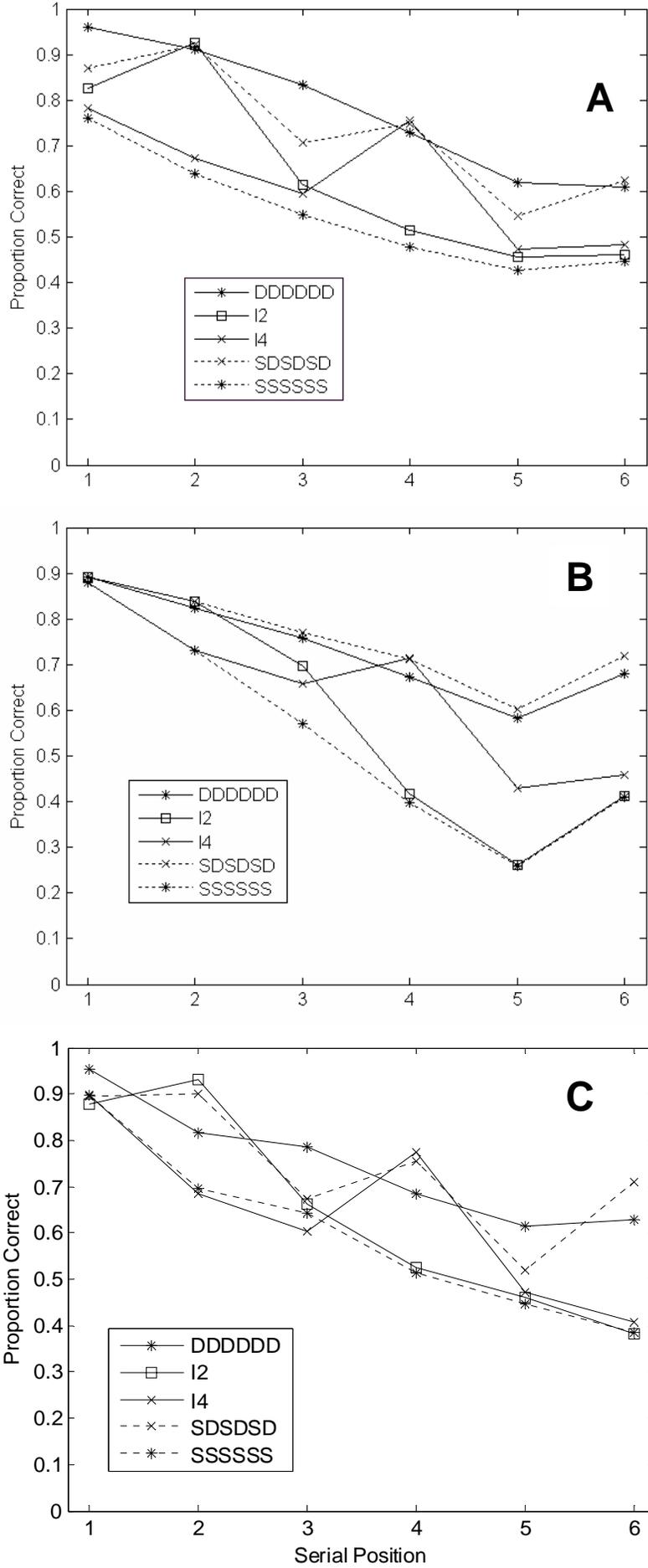


Figure 5

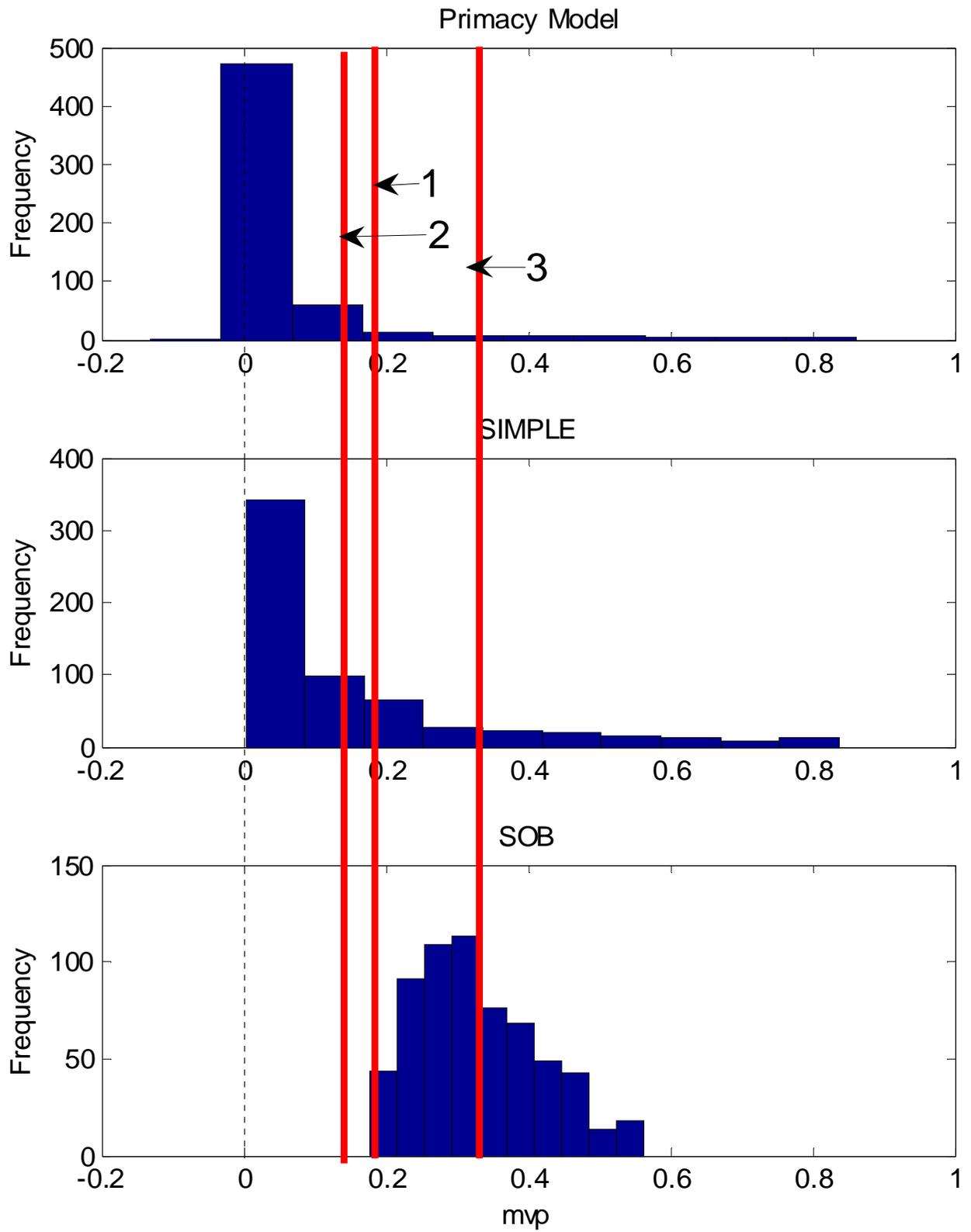


Figure 6

