Modeling Verbal Short-Term Memory: A Walk Around the Neighborhood

Jean Saint-Aubin

Université de Moncton

Marie Poirier

City, University of London

James M. Yearsley

City, University of London

Jean-Michel Robichaud

Université de Moncton

Dominic Guitard

Cardiff University

Authors' Note

This research was supported by a discovery grant from the Natural Sciences and Engineering Research Council of Canada to Jean Saint-Aubin. While working on this article, Dominic Guitard was supported by a postdoctoral fellowship from NSERC and Jean-Michel Robichaud was supported by an NSERC Undergraduate Student Research Award. We thank Peter Graf for suggesting Experiment 2 at a conference. Correspondence concerning this article should be addressed to Jean Saint-Aubin, École de psychologie, Université de Moncton, Moncton, New Brunswick, E1A 3E9, Canada.
Contact: Jean.Saint-Aubin@umoncton.ca

Open Practices Statement

The data are available on the Open Science Framework project page (osf.io/ga49s).

Abstract

When remembering over the short-term, long-term knowledge has a large effect on the number of correctly recalled items and little impact on memory for order. This is true for example, when the effects of semantic category are examined. Contrary to what these findings suggest, Poirier et al., (2015) proposed that memory for order relies on the level of activation within long-term networks. Importantly, although their view has been criticized, they showed that manipulating semantic associations led to item migrations that were atypical. In this paper, we show that similar migrations can be obtained with another knowledge-based factor: orthographic neighborhood. In three experiments, we manipulated the orthographic neighborhood of to-be-recalled items. The latter is a sublexical factor; as such, it is much less likely than semantic relatedness to involve demand characteristics or grouping strategies. The first experiment established that the neighborhood manipulation produced the pattern of item migrations previously observed with semantic relatedness, confirming that the migration effect can generalize to other variables. The last two experiments suggested that migrations were due to the features shared across list items rather than to item co-activation (as in Poirier et al.). The results were successfully modelled by calling upon the Revised Feature Model (Saint-Aubin et al., 2021), where recall depends on selecting a retrieval candidate based on the features of the cueing information. Overall, our findings underline the usefulness of a model where retrieval is determined by relative distinctiveness and underlines that multiple mechanisms can lead to order errors in recall.

Keywords: Short-term memory; working memory; order recall; immediate memory; Revised Feature Model

Modeling Verbal Short-Term Memory: A Walk Around the Neighborhood

Consider the following: you decide to use your phone to clear your credit card bill. You complete a few operations and receive a 6-digit code sent by the bank. You glance at it, noticing that the last three digits reproduce an old address. You then return to your banking application and input the sequence correctly; bill paid. How did you remember the 6-digit sequence in order? What effect did familiarity with part of the sequence have? Here, we explore such questions by examining the interplay between prior knowledge and verbal short-term memory (STM).

STM is thought of as carrying out the temporary maintenance necessary for many cognitive tasks (e.g. Baddeley, 1986; Camos, 2015). It is thought important in everyday cognition (Cowan, 1999; Engle, 2018) and is regarded as central in maintaining order information (Majerus, 2009). Consider the example provided above; the order of the sequence is critical. The same is true in several everyday tasks such as remembering the order of the three things you set out to do when leaving your desk, remembering the next few steps in a recipe, the brief series of instructions from a teacher or seminar leader, or when learning a new multiple syllable word or name.

The study of verbal STM–and of its interaction with long-term memory (LTM)—have often relied on immediate serial recall, where participants attempt to remember a small number of items, in order, just after their presentation. Long-term knowledge of the language significantly impacts performance of this task. Word frequency enhances performance (Poirier & Saint-Aubin, 1996), as do concreteness (Guérard & Saint-Aubin, 2012), lexicality (Hulme et al., 1991; Saint-Aubin & Poirier, 2000) and category membership (see Neath, Saint-Aubin & Surprenant, 2022; Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999b). The same is true for sub-lexical factors; for example, non-words containing more familiar syllables are better recalled (Thorn & Frankish, 2005). This influence of LTM is expected by models that suggest considerable overlap between STM and the semantic, lexical, and sublexical networks thought to underlie language representation (e.g., Acheson & MacDonald, 2009; Kowialiewski et al., 2021; Majerus et al., 2004). As STM is thought to involve these long-term networks, in turn, the characteristics of these networks should influence performance.

For example, Roodenrys (2009) argued that to explain the impact of several lexical and sub-lexical variables on STM, one needs an interactive network model which includes letter, phonemic and lexical levels of representation. Other models include semantic levels of representation also (e.g., Buchsbaum & D'Esposito, 2008; Cowan, 1999; Cowan & Chen, 2009; Gupta, 2009; LaRocque et al., 2014; Majerus, 2009; Martin & Gupta, 2004; R.C. Martin, 2006; Acheson, et al.,2011). However, there remains much debate and uncertainty as to the manner in which LTM and STM are integrated (e.g. Cowan, 2019; Kowialiewski, et al. 2021; Norris, 2017, 2019).

This paper aims to contribute to this important area. To do so, we focus on a new empirical effect first reported by Poirier et al. (2015). These authors showed that sematic relatedness between the items in an STM task could lead to item migrations during recall that violate the typically observed patterns. We will return to the details of these findings later on. In Experiment 1, we asked if this novel finding could be reproduced by manipulating another LTM variable not based on semantic representations, namely orthographic neighborhood. Words are orthographic neighbors when they share all their letters, in the same positions, except one. For instance, the orthographic neighbors of ***bat*** include *eat, but, oat,* and *bay*.

In essence, we tested the generality of the migration effect reported by Poirier et al. (2015) – i.e. are the findings specific to semantic relatedness, or is the pattern apparent when items share other, less obvious, LTM features? Orthographic neighborhood is a sublexical variable, where shared features are not related to meaning. This makes the relatedness of the items much less obvious, with the implication that demand characteristics and grouping strategies are unlikely.

Prior research has shown that words with many neighbors are better recalled than words with less dense neighborhoods (Derraugh et al., 2017; Roodenrys, 2009; Roodenrys et al., 2002). Furthermore, the effect is not driven by response modality, since it has been observed when participants recalled the items aloud, typed them or clicked on them (Derraugh et al., 2017; Roodenrys et al., 2002). The typical interpretation of this effect is that neighbors co-activate each other within the lexical network, leading to better recall of words with larger neighborhoods (Roodenrys, 2009).

Activation and Order Recall

Poirier et al. (2015) suggested that manipulating relative item activation would lead to predictable STM effects on order recall. They proposed a model that included a primacy gradient where items are encoded with a decreasing level of LTM activation from the first item onward (see Page & Norris,1998, for a related proposal, which, however, does not involve LTM). At retrieval, a noisy process selects items based on activation level, which allows the order of the items to be reconstructed, albeit with some error. Poirier et al. called the framework the "Activated Network" view or ANet. ANet predicts that if an item's activation is *increased*, it will tend to be recalled earlier in the list. Because of the proposed primacy gradient, increasing the (relative) activation of an item appearing later in a list would make it more like earlier items, increasing the chances the item will be recalled early, out of order.

To test this idea, Poirier et al. (2015) had a target item appearing in position 5, within 6-item lists, where the first three items were strong associates of the target (e.g. *band, record,* *concert*, yellow, ***music***, tourist). According to ANet, the first three items would increase the activation of the target (5th item), making it more like earlier items; the prediction was that the target would migrate towards earlier positions more often than controls. Results supported this prediction. Another experiment compared lists where all the items were from the same semantic category with lists of unrelated words. Again, the critical words showed more frequent migrations to earlier positions when the surrounding items were expected to heighten their activation. However, using associates in the first three positions could lead to a grouping strategy that is unrelated to activation and might also involve demand characteristics; this is less true of the study using semantic category – but ANet predictions clearly requires further and stronger tests.

Recently, Kowialiewski et al. (2021) tested a connectionist model of STM for order based on the primacy gradient idea included in the Poirier et al.'s (2015) proposal; they concluded that on its own, an LTM activation gradient could not account for the interaction between LTM factors and STM order memory. Also, the Poirier et al. proposal suggests that to-be-remembered items are represented through LTM activation. On the basis of a number of arguments, Norris (2017, 2019) has insisted that there has to be a representation of recently presented items that is separate from prior LTM knowledge or prior LTM representations. Hence, there are significant challenges to the idea of an activation-based primacy gradient. Additionally, the ANet model suggests that activation at multiple levels within the network should have an impact – but only semantic / meaning-based associations have been studied. For the ANet proposal to warrant revision or review, given the challenges mentioned above, activation effects from other network levels must be demonstrated.

We pursued these aims in the work reported here; we used the associations between orthographic neighbors to manipulate activation as the latter provides a stronger test of ANet predictions. More specifically, we called upon three experiments to achieve two main objectives; the first was to attempt to extend the migration effect reported by Poirier et al. (2015) to another variable, namely orthographic neighborhood. The second was to assess two hypotheses to account for the migration effect. The first hypothesis is embodied in ANet and suggests that increased activation of a target induced by the presentation of three orthographic neighbors within the same list will produce migrations towards earlier serial positions. The second hypothesis is based on an alternative proposal, expressed in a very different model, namely the Revised Feature Model (RFM; Saint-Aubin et al., 2021). In the latter model, migrations are attributed to similarity between the features of the orthographic neighbors within a list. Said view does not rely on network activation as an explanatory concept. Instead, the RFM highlights the fact that semantically related and neighboring items can be thought of as sharing a number of distinct features – relative to items that are not part of the same neighborhood. Within this perspective, if a retrieval cue shares a number of features with multiple retrieval candidates, then the probability of retrieval is increased, although said retrieval might generate an order error.

In Experiment 1, we used a strategy like that used in the first experiment of Poirier et al. (2015), except that instead of manipulating semantic relatedness, we called upon neighborhood. In the experimental lists, the first three items were neighbors of the fifth item presented. ANet and the RFM predict that the fifth item would migrate forward. According to ANet the migration would be due to the additional activation of the fifth item by the three related items, while according to the RFM the migration would occur because of the shared features. In Experiment 2, we replicated the design of Experiment 1 and importantly, we added a condition in which the last three items were neighbors of the third item. This second experiment allowed us to contrast the opposing predictions of ANet and the RFM.

To anticipate, the predictions derived from ANet were not supported, while the RFM made the correct predictions and could be called upon to model the findings. Before further considering explanatory proposals of the migration effect, Experiment 3 asked if the relevant factor in the impact of orthographic neighborhood was the increase on phonological similarity of the neighbors, of if the orthographic structure was indeed the driver of the observed effects.

Experiment 1

Method

Participants. Thirty-six undergraduate students from the Université de Moncton volunteered to participate in exchange for a small honorarium ($5). All participants reported normal or corrected to normal vision and were native French speakers. Informed consent was obtained from all participants prior to participation and the study was approved by the university's research review committee.

Sample size. The sample size calculation was based on the critical interaction between the conditions and serial positions for order errors of the critical item in Position 5 of Poirier et al.’s (2015) first experiment as the current experiment was modeled after their experiment. We used G*Power (Faul et al., 2009) to guide the selection of our sample size. The size of the interaction between condition and serial positions in Poirier et al. was found to be $f = 0.265$. Using this effect size, the results from an a priori power analysis with an alpha of .05, a power of .95 and default parameters for the correlation between the repeated measures and the non-sphericity correction, revealed that 28 participants would allow us to detect such an interaction. However, because the size of the neighborhood effect is unknown, we decided to overpower our design. Sensitivity analysis with the same parameters revealed that 36 participants would allow us to detect a smaller interaction $f = 0.23$.

Materials. The experiment included 36 lists; each participant was presented with 18 experimental and 18 control lists. We first selected 36 French target words along with three neighbors for each using Lexique (New et al., 2004). We then created two sets of 18 lists in which the target word was placed in the 5^{th} position and the corresponding neighbors were placed in Position 1, 2, and 3; the remaining positions (4, 6 and 7) were filled with unrelated words. One half of the experimental lists were randomly chosen to become Set 1, while the other half were assigned to Set 2. The lists from each set were then used again to create two further sets of 18 control lists. More specifically, the control lists had the same three associated words in the first positions, in the same order. However, the last four words were a quasi-random selection (without replacement) from the remaining words of each set. The only constraints were that the 5^{th} item was not a neighbor of those in Position 1 to 3 and that the 5^{th} item was not semantically related to the items in Position 1, 2 and 3 in any obvious way. An example is provided in Table 1. For each participant, there was a random selection of the experimental set of 18 lists; the only constraint being that across participants, each set was used as often. If a participant was allocated Set 1 for the experimental trials, then s/he was given Set 2 for the control lists. In this way, items were never repeated for a given participant and the content of the experimental and control lists was counterbalanced across participants. To be clear, each participant studied each word once, for half the participants the words used in the control list became the experimental list for the other half. All words were presented in black, lowercase, 28-point Arial font, at the center of a 47.72 cm (18 inches) computer screen with a resolution of 800 X 600 pixels. The experiment was controlled with E-Prime 2.0.

Design and Procedure. A 2 x 7 repeated-measures design was implemented with list type (control, experimental) and serial position (1 to 7) as factors. The order of the 36 trials was randomized independently for each participant and they were preceded by four practice trials. Participants were tested individually in one experimental session lasting approximately 30 minutes.

Each trial began with the presentation of a white screen for 1000 ms followed by the presentation of the seven to-be-remembered words that were displayed at a rate of one word per second (1,000 ms on, 0 ms off) at the center of the screen. After the presentation of the last word, three questions marks, “???” in blue appeared in the upper middle part of the screen as a cue for participants to recall the seven words they had just seen. Participants were instructed to recall the words from the first to the last word that had been presented. They were to type-in their responses; as soon as they began typing, the letters appeared centered at the bottom of the screen. Participants pressed the space bar to type their next response. Immediately upon pressing the space bar, the word disappeared from the screen, and participants could begin typing their next response. Participants were instructed to type an “x” whenever they did not know the word at a given serial position. Once an item or an “x” had been typed, the participant pressed the space bar to register the response. There was no time limit for recall and participants were not allowed to backtrack to change previous responses. Participants initiated the next trial by pressing enter.

Results

As mentioned above, the critical item and its control were in the fifth serial position of their respective lists. We first computed the proportion of correct recall with a strict serial recall criterion. With this criterion, an item ought to be recalled at its presentation position to be scored correct. We also computed the proportion of items recalled with a free recall criterion; in this case, a recalled item was scored correct irrespective of position. As shown in Table 2, with both strict and free recall criteria, overall mean performance averaged across positions was very similar between the control and experimental conditions when omitting the critical position 5, t (35) = 1.17, p =.248, d = 0.20 and t (35) = 1.24, p =.223, d = 0.21, respectively. A different picture emerged for Item 5 which was better recalled in the experimental than in the control condition with both a strict, t (35) = 2.52, p = .016 d = 0.42, and a free recall criterion, t (35) = 4.50, p <.001, d = 0.75.

Item 5. We then analyzed the pattern of migrations of the fifth item. As shown in Figure 1, the fifth item is much more likely to be erroneously recalled in the first three serial positions if it was part of an experimental than of a control list. There are no further clear differences at the other serial positions. These observations were tested with a 2 (condition) x 6 (error position) repeated measures ANOVA. Results revealed that errors were significantly more frequent for the experimental than for the control condition, F (1, 35) = 19.22, p <.001, η_p^2 = .36. There was also a main effect of position, F (5, 175) = 6.69, p <.001, η_p^2 = .16, and as predicted, a significant interaction, F (5, 175) = 4.98, p <.001, η_p^2 = .13. Post hoc Tukey's HSD tests showed that Item 5 migrated more often to Position 1 (p <.001), 2 (p <.001), and 3 (p = .058) in the experimental than in the control condition, while there was no evidence of more migrations to the other positions (all ps>.29).

Discussion

The critical item in Position 5 was—overall—more likely to be recalled in the experimental than in the control condition (Derraugh et al., 2017; Poirier et al., 2015). This finding suggests that the activation of this item was successfully increased by the inclusion of three of its neighbors in earlier list positions, as expected based on the typical interpretation of the neighborhood effect (Derraugh et al., 2017). Critically, in addition to being more frequently recalled, the target item in the experimental condition migrated forward, i.e. more toward the beginning of the list. Those migrations were predicted because of the activation-based view proposed by ANet. In a nutshell, the primacy gradient means a decreasing level of activation from the first item onwards and recall is assumed to be based on activation levels. This means the first item is the most likely to be recalled in the first position, the second item is then the most likely to be recalled, and so on. Because of the heightened activation of the 5^{th} item, it is assumed that it will be in a state that is much more similar to the earlier items than it would otherwise be. The implication is that relative to the equivalent 5^{th} item in the control list, significantly more forward migrations of the target item were expected (see Figure 1). In effect, the increase in activation appears to have been sufficient for the target 5^{th} item to have its migration peak move to Position 2, away from the usual peak around Position 4. The more typical migration pattern is seen for the Control condition in Figure 1. It is worth noting that the critical item was *less likely* to move to Position 4 than the control item, although the difference was not significant.

Experiment 2

Although broadly supporting the predictions derived from ANet, results of Experiment 1 can also be accounted for by a feature-based account. In effect, because Item 5 is an orthographic neighbor of the first three items, it shares more orthographic and phonological features with the first items. Due to those shared features, Item 5 would be more likely to be erroneously recall as one of the first three items. To contrast the predictions derived from ANet with those based on a putative feature-based account, we introduced two critical conditions. The first one is the same as in Experiment 1 with the critical Item 5 related to the first three items. The inclusion of the first condition allows us to achieve two aims. First, before choosing between two theories, it is essential to establish that a novel phenomenon is reproductible (Simons, 2014). Second, in Experiment 1, the impact of our manipulation on the migration of the critical Item 5 to the fourth position when it is related to the first three items was unclear. This effect is central for choosing between ANet and a feature-based interpretation. According to ANet, the target item should have moved to Position 4 more frequently when it is related to the first three, while the descriptive statistics show a trend in the opposite direction. However, this pattern fits well with a feature-based account. Unfortunately, possibly due to a lack of power, the trend on Position 4 failed to reach significance in Experiment 1. Therefore, 100 participants took part in Experiment 2, instead of 36 as in Experiment 1. The sensitivity analysis computed with G*Power revealed that 100 participants would have a power greater than .95 to detect an effect at Position 4 as small as Cohen's d = 0.36 (Faul et al., 2009). In addition, to anticipate, we modeled data of Experiment 2 with the RFM and the selection of 100 participants allowed the achievement of more stable estimates for the simulations.

In the second critical condition, the target item was located in Position 3 and related to the last three items. According to ANet, the critical item should move to the first two serial positions albeit a ceiling effect might prevent seeing the effect on the first serial position. On the other hand, according to a feature-based account, the critical item should move more frequently to the last three serial positions.

Method

Participants. One hundred volunteers from an online data collection agency, Prolific (https://www.prolific.co/), participated and were paid £2.50. Inclusion criteria for this study were as follows: (a) The participant must be a native French speaker; (b) their nationality must French, Belgium, Swiss, or Canadian; (c) they must have normal or corrected-to-normal vision; (d) they must have no cognitive impairment or dementia; (e) they must have no language-related disorders; (e) their age must be between 18 and 30 years; and (f) they must have an approval rating of at least 90% on prior submissions at Prolific. Inclusion criteria (a) through (e) were self-reported, and the approval rate is objectively computed by Prolific. Informed consent was obtained from all participants prior to participation and the study was approved by the university's research review committee.

Materials. The materials were identical to Experiment 1 except for the following changes. The experiment included 36 lists; each participant was presented with 18 experimental (9 experimental lists 3^{rd} position, 9 experimental lists 5^{th} position) and 18 control lists (9 control lists 3^{rd} position, 9 control lists 5^{th} position). We created two sets of 18 lists in which the target word was placed in the 3^{rd} position and the corresponding neighbors were placed in Position 5, 6, and 7; the remaining positions (1, 2 and 4) were filled with unrelated words. We also took the two sets of 18 lists form Experiment 1 in which the target word was placed in the 5^{th} position and the corresponding neighbors were placed in Position 1, 2, and 3; the remaining positions (4, 6 and 7) were filled with unrelated words. One half of the experimental lists were randomly chosen to become Set 1, while the other half was assigned to Set 2. The lists from each set were then used again to create two further sets of 18 control lists for the target word in the 3^{rd} position and two further sets of 18 control lists for the target word in the 5^{th} position. More specifically, the control lists for the 3^{rd} position had the same three associated words in the last positions while the control lists for the 5^{th} position had the same three associated words in the first positions, in the same order for both cases. However, for both 3^{rd} and 5^{th} position control lists, the remaining four words were the unrelated remaining words of each set. The only constraints were that the 3^{rd}

item was not a neighbor of those in Positions 5 to 7 for the control lists of the 3rd item and that the 3rd item was not semantically related to the items in Position 5 to 7 in any obvious way. The same criteria were applied for the 5th item except that latter item was not a neighbor of those in Positions 1 to 3 and that the 5th item was not semantically related to the items in Position 1 to 3 in any obvious way. Across participants, each set was used as often, and items were never repeated for a given participant and the content of the experimental and control lists was counterbalanced across participants. All words were presented in white, lowercase, 30-point Times New Roman font, at the center of the computer screen on a black background. The experiment was controlled with PsyToolKit (Stoet, 2010, 2017).

Design and Procedure. A 4 x 7 repeated-measures design was implemented with the list type (control [item 3 or item 5], experimental [target item in position 3 or 5]) and serial position (1 to 7) as factors. The order of the 36 trials was randomized independently for each participant and they were preceded by four practice trials. Participants were tested individually in one experimental session lasting approximately 20 minutes. The other elements of the design and the procedure were identical to Experiment 1, except for the following changes. Each trial began with the presentation of a black screen instead of a white screen for 1000 ms and participants initiated the next trial by pressing the space bar instead of enter key.

Results

As mentioned previously, on half the experimental lists, items in Position 1, 2, 3, were orthographic neighbors of the target item which was in Position 5, and on the other half of the experimental lists, items in Position 5, 6, 7, were orthographic neighbors of the target item which was in Position 3. In line with Experiment 1, performance averaged across positions was examined with both strict and free recall criteria. We examined performance separately for trials in which the critical item was in Position 3 and Position 5.

Position 3. When omitting the critical item in Position 3, performance averaged across positions did not differ between the control and the experimental conditions with a strict serial recall criterion, $t < 1$, $p = .691$, $d = 0.04$, or a free recall criterion, $t < 1$, $p = .500$, $d = 0.07$. For the critical item in Position 3, with a strict serial recall criterion, performance did not differ between the experimental and the control condition, $t(99) = -1.31$, $p = .195$, $d = 0.13$. However, with the free recall criterion, participants' performance was superior in the experimental than in the control condition, $t(99) = 3.23$, $p = .002$, $d = 0.32$.

Position 5. When omitting the critical item in Position 5, recall performance did not differ between the control and the experimental condition with a strict serial recall criterion, $t < 1$, $p = .410$, $d = 0.08$, or a free recall criterion, $t(99) = -1.18$, $p = .241$, $d = 0.12$. For the critical item in Position 5, with a strict serial recall criterion, performance was similar for the control and the experimental condition, $t(99) = 1.30$, $p = .195$, $d = 0.13$. However, with a free recall criterion, participants were better in the experimental than in the control condition, $t(99) = 4.82$, $p < .001$, $d = 0.48$.

As shown in Figure 2, as in Experiment 1, when the target item was in Position 5, it was more frequently recalled at an earlier serial position than its control counterpart. However, when the target item was in Position 3, it was more frequently recalled at later serial positions than its control counterpart. Migrations were analyzed separately for each target position.

Item 3. The proportion of trials for which the critical item (in Position 3) in the experimental condition was recalled at another position was compared to the control condition with a 2 (condition) x 6 (error position: 1,2,4,5,6,7) repeated measures ANOVA. Item 3 migrated more frequently in the experimental than in the control condition, F (1, 99) = 7.90, p =.006, η_p^2=.07. There was also a main effect of position, F (5, 495) = 8.98, p <.001, η_p^2=.08, but the interaction did not reach conventional level of significance, F (5, 495) = 1.80, p =.111, η_p^2=.02. Given the theoretical importance of the interaction, we computed Post hoc Tukey's HSD tests which revealed that Item 3 migrated more often to the Position 5 (p = .007) and Position 6 (p = .044) in the experimental than in the control lists—while the patterns of migration did not differ for the other positions (all ps > .10).

Item 5. The proportion of trials for which the critical item (in Position 5) in the experimental condition was recalled at another position was compared to the control condition with a 2 (condition) x 6 (error position: 1,2,3,4,6,7) repeated measures ANOVA. Like Experiment 1, the ANOVA revealed a main effect of condition with more errors in the experimental than in the control condition, F (1, 99) = 32.53, p < .001, η_p^2 = .25, a main effect of serial position, F (5, 495) = 20.48, p <.001, η_p^2 = .17, and the expected interaction, F (5, 495) = 2.78, p = .017, η_p^2 = .03. Post hoc Tukey's HSD tests showed that there were more migrations of Item 5 in the experimental than in the control lists at Position 1 (p < .001), Position 2 (p < .001) and Position 3 (p = .003). There was no difference at Position 4 (p = .644) or the other positions (all ps > .35).

Discussion

Results of Experiment 2 with the target item in Position 5 nicely replicated those observed in Experiment 1 and those observed with semantic relatedness by Poirier et al. (2015). It is worth mentioning that contrary to previous experiments, Experiment 2 took place online with a sample comprising participants from four different countries. Therefore, although strictly speaking results are not coming from different laboratories, it can be said that they partly meet the reproducibility criterion for including a phenomenon in the list of benchmark findings (Oberauer et al., 2018). Using a more powerful design with three times more participants than in Experiment 1, results provide a clear answer about migrations to Position 4. Contrary to ANet predictions, the target item in Position 5 did not migrate more frequently to Position 4 in the experimental than in the control condition.

The condition with the target in Position 3 was the most critical for adjudicating between ANet and a feature-based explanation. According to ANet, in the critical condition, the target item should have migrated more toward Position 1 and 2, while the feature-based explanation predicted the reverse with migrations toward Position 5, 6 and 7. Results were clear with more migrations in Position 5 and 6.

Experiment 3

Despite the clarity of the observed results, it remains to be seen whether they are due to orthographic or phonological features. Orthographic neighbors naturally tend to share phonemes, therefore introducing phonological similarity among them. Nevertheless, we believe our results are driven by orthographic neighborhood rather than phonological similarity. In effect, the presence of orthographic neighbors improves item recall of the target, while phonological similarity has the opposing effect (Roodenrys, Guitard et al., 2022; Roodenrys, Miller et al., 2022). Furthermore, it is worth noting that French is an opaque language in which the mapping from orthography to phonology is weak (Seymour et al., 2003). In our materials, orthographic neighbors were not automatically phonologically similar. This is illustrated by the phonological transcription of the following experimental list from Experiment 1: **bave /bav/ – taie /tɛ /– brie / bʁi /**– œil /ɔεj/– **baie /bɛ /**– chou / ʃu/– vœu / vɔø/. However, it is true that there were some phonological similarities among them.

To further establish that our results were driven by orthographic neighborhood rather than phonological similarity, we created a new set of stimuli in which the target item shares no phoneme with its neighbors, although the latter can share some phonemes together. Under the assumption that our results are due to the confounded effect of phonological similarity a new pattern of results should emerge. Conversely, if as we assume, orthographic neighborhood is the key factor producing our results, we should replicate them.

Method

Participants. Two hundred volunteers from Prolific (https://www.prolific.co/) who did not participate in the previous experiments took part in this experiment and were paid £1.25. The inclusion criteria were identical to those used in Experiment 2. Because of the limited number of available word pools in which the target did not share any phoneme with its orthographic neighbors, target location (Position 3 or 5) was a between-participants factor instead of a within-participant factor. Therefore, the overall number of participants was twice the number used in Experiment 3 to achieve the same number of observations per condition.

Materials. The materials were similar to Experiment 2 except for the following critical changes. In this experiment, participants were presented with 6 experimental lists and 6 control lists. Half of the participants were assigned to the 3rd position manipulation and the other half were assigned to the 5th position manipulation. For participants in the 3rd position manipulation, two sets of 12 lists (6 control lists, 6 experimental lists) were created in which the target word was located in the 3rd position and the corresponding neighbors were located in Position 5, 6, and 7; the remaining positions (1, 2 and 4) were filled with unrelated words. Likewise, for participants in the 5th position manipulation, two sets of 12 lists (6 control lists, 6 experimental lists) were created in which the target word was located in the 5th position and the corresponding neighbors were located in Position 1, 2, and 3; the remaining positions (4, 6 and 7) were filled with unrelated words. Across participants, for both the participants in the 3rd position manipulation and the participants in the 5th position manipulation, each set was used equally often, items were never repeated for a given participant within an experimental session and the words used in the control and experimental lists were counterbalanced across participants.

In this experiment, the key difference is the great care taken to ensure that the target item was orthographically similar to the three related items, but totally phonologically dissimilar. As shown in Table 1, related words could be phonologically similar to one another but could never share a phoneme with the target word. This was accomplished by taking advantage of inconsistencies between phonology and orthography in the French language.

Design and Procedure. A 2 x 2 x 7 mixed design was implemented with list type (control vs. experimental) and serial position (1 to 7) as repeated measure factors and item (item 3 vs. item 5) as a between-participants factor. The order of the 12 experimental trials was randomized independently for each participant and they were preceded by four practice trials. Half of the participants were assigned to the 3rd position manipulation and the other half were assigned to the 5th position manipulation. All participants were tested individually in one online experimental session lasting approximately 10 minutes. All other details were identical to Experiment 2.

Results

As in previous experiments, performance averaged across positions was examined with both strict and free recall criteria. We examined performance separately for trials in which the critical item was in Position 3 and Position 5.

Position 3. As shown in Table 2, when omitting the critical item in Position 3, performance averaged across positions did not differ between the control and the experimental conditions with a strict serial recall criterion, t (99) = -1.06, p =.290, d = 0.11, or a free recall criterion, t(99) = -1.63, p =.107, d = 0.16. However, for the critical item in Position 3, recall performance was superior in the experimental than in the control condition with a strict serial recall criterion, t (99) = 2.63, p =.01, d = 0.26, and a free recall criterion, t (99) = 3.64, p <.001, d = 0.36.

Position 5. When omitting the critical item in Position 5, an examination of Table 2 reveals that recall performance did not differ between the control and the experimental condition with a strict serial recall criterion, t -1.63, p =.107, d = 0.16. However, with a free recall criterion, performance was better in the experimental than in the control condition, t (99) = 2.41, p =.018, d = 0.24. For the critical item in Position 5, performance was superior in the experimental than in the control condition with a strict serial recall criterion, t (99) = 2.36, p =.021, d = 0.24, and a free recall criterion, t (99) = 3.62, p < .001, d = 0.36.

An inspection of Figure 3 revealed that the target item in Position 5 was more frequently recalled at an earlier serial position than its control counterpart, whereas the target item in Position 3 was more frequently recalled at a later serial position than its control counterpart. This pattern nicely reproduced the pattern observed in the first two experiments. As in Experiment 2, migrations were analyzed separately for each target position.

Item 3. In this section we replicated the analyses of Experiment 2 for the critical item in Position 3. Results from the analysis revealed that Item 3 did not migrate more frequently in the experimental than in the control condition, F (1, 99) = 3.77, p =.055, η_p^2=.04. There was a main effect of position, F (5, 495) = 6.453, p <.001, η_p^2=.06, but the interaction did not reach the conventional level of significance, F (5, 495) = 1.307, p =. 260, η_p^2=.01. However, given the theoretical importance of the interaction, we computed Post hoc Tukey's HSD tests which revealed that Item 3 migrated more often to the Position 5 (p = .032), but failed to reach the conventional level of significance for Position 6 (p = .072), and Position 7 (p = .091) in the experimental than in the control lists. The pattern of migration did not differ for the other positions (1, 2, 4), all ps > .60.

Item 5. In this section we replicated the analyses of Experiment 2 for the critical item in Position 5. Results from the analyses revealed that all main effects and the two-way interaction reached conventional level of significance. More exactly, participants made more errors in the experimental than in the control condition, F (1, 99) = 7.61, p = .007, η_p^2 = .07, there was a main effect of serial position, F (5, 495) = 13.62, p <.001, η_p^2 = .12, and a two-way interaction between condition and serial position, F (5, 495) = 3.31, p = .006, η_p^2 = .03. Post hoc Tukey's HSD tests showed that there were more migrations of Item 5 in the experimental than in the control lists at Position 1 (p = .016), Position 2 (p < .001) and Position 3 (p = .006). There was no difference at Position 4 (p = .622) or the other positions (all ps > .34).

Discussion

Results of Experiment 3 nicely replicated those observed in Experiment 2, while removing the confounded influence of phonological similarity. As such, our results provide additional evidence that orthographic neighborhood per se influences short-term ordered recall performance (Derraugh et al., 2017). Furthermore, by reproducing the results of Experiment 2 with a new set of stimuli, we demonstrated that the effect is reproductible and is not due to the specificity of the stimuli used in the first two experiments (see, e.g., Guitard et al., 2018; Neath et al., 2021).

Modelling similarity-based migrations

The results of Experiments 2 and 3 seem to pose a challenge for ANet, but leave open the possibility of a feature-based account of these effects. However, it is important to confirm this intuition by attempting to quantitatively fit these results using a specific feature-based model of encoding and recall, since intuition can be a poor guide when dealing with accounts of complex phenomena (Guest & Martin, 2021). The specific model we shall use is the Revised Feature Model (RFM), which has recently been called upon to account for a number of empirical phenomena, in both the verbal and visuo-spatial domains (Poirier et al., 2019) and in both short- and long-term memory tasks (Cyr et al., 2021; Saint-Aubin et al., 2021). The RFM is built on the same similarity-based architecture as the original Feature Model (Nairne, 1990; Neath & Nairne, 1995; Neath & Surprenant, 2007), but includes a rehearsal mechanism and other small updates which allow the model to provide a quantitative fit to experimental data.

Full details of the model can be found in Cyr et al. (2021) and Saint-Aubin et al. (2021), and the code used to implement the model is available on the Open Science Framework page; here we provide a brief overview of the main features. In the RFM, items are represented by vectors of randomly set features that, usually, take values 1-3, or 0 for a feature which has been overwritten, about which more below. In typical applications we assume items such as the ones presented in these experiments have 20 modality independent features representing the content of the item, and a further two modality dependent features that encode details relating to the presentation format. The RFM assumes information about items is stored in two places; a secondary memory store which retains full information about the presented items, and a primary memory store which stores copies of the items which are to be used as cues to recall the full items from secondary memory. The cues stored in primary memory are gradually degraded by a process of retroactive interference – if an item *n* is presented after an item *m*, and a particular feature of item *n*, $f^n{}_i$ matches the same feature in item *m*, $f^m{}_i$, then this feature of item *m* will be overwritten (i.e. set to 0) with probability $e^{-\lambda(n-m+1)}$, where λ is some positive constant. Under the right conditions rehearsal after presentation of an item q can act to restore overwritten features of all previous items, with a probability given by $re^{-\frac{(q-1)^2}{9}}$, where $0 \le r \le 1$ controls the effectiveness of rehearsal, and the exponential function has the effect of suppressing the effectiveness of rehearsal once the number of presented items exceeds four (e.g. Bhatarah et al., 2009). After list presentation, there is a final process of overwriting of modality-independent features due to continuing internal thought activity in preparation for recall.

In addition to features relating to the item, the RFM also assumes that each to-be-remembered item or cue is given a positional code determining where in the list it was presented. Between presentation and recall these codes can drift, in a manner similar to that introduced by Estes (1972).

At recall, the model first picks the position of the item to be recalled, i, and from that identifies the relevant cue c_i (errors may be introduced at this stage due to drift in the positional coding.). The cue c_i, which is the degraded representation of the item i in primary memory, is used to identify the to-be-recalled item via the similarity between this cue and all traces in secondary memory [that is, the items included in the most recent list]. The similarities are related to the feature-to-feature correspondence between cue $S_{ij} = e^{-a\,p_{ij}}$ where p_{ij} is the proportion of mis-matching features between cue i and trace j, and $a > 0$ is a constant. The probability of retrieving item j given cue i, is then given by,

$$p(j,i) = \frac{e^{\frac{S_{ij}}{\tau}}}{\sum e^{\frac{S_{ik}}{\tau}} + e^{\frac{S_0}{\tau}}}, \qquad p(0,i) = \frac{e^{\frac{S_0}{\tau}}}{\sum e^{\frac{S_{ik}}{\tau}} + e^{\frac{S_0}{\tau}}}$$

Here recalling item ‘0’ results in an omission, τ is a temperature parameter that controls how deterministically the item with the highest similarity to the cue is recalled, and S_0 is a constant which can be thought of as the minimum cue-item similarity needed to reliably generate a recall. The model also includes a step where multiple recalls of the same item are suppressed, and instead result in an omission. One small change from previous implementations of the RFM, which proved necessary to fit the combination of serial positions, error gradients, and omission rates, is that any time an omission occurs, there is a fixed (50%) probability that list recall will terminate, and further items will be automatically omitted.

In all our lists, items 1 to 3 or 5 to 7 are orthographic neighbors of a target item, which is a member of the list in the experimental condition, but omitted in the control condition. If two items are each orthographic neighbors of a third item then they need not be orthographic neighbors of each other, but they will nevertheless share some letters. There is therefore a particular similarity structure in the lists, whereby items 1-3 or 5-7 are both similar to some target item, and somewhat similar to each other. Note that this similarity is orthographic; otherwise, items will be quite distinct – i.e. meaning, category, characteristics, etc.

In the language of a feature-based model similarity or dissimilarity between items equates to matching or mismatching features, since similarity is computed on the basis of proportion of matching features. Therefore, in the control condition, for example, we need to ensure items 1-3 or 5-7 share the values of some subset of features. In the experimental condition each of these items needs to share at least that many features with the target item. Note that although the vector of features is supposed to encode all the information about the item, the encoding is rather abstract in the RFM, so that it is not really possible to identify which value of which feature corresponds to the first letter of a word being '*t*' for example.

To handle the orthographic similarity between items 1-3 or 5-7 we therefore started by generating a target item, and then fixing six of the 20 modality independent features of items 1-3 or 5-7 to take the same value as those of the target item. The location of four of these matching features was fixed for items 1-3 or 5-7, while the location of the other two features varied. As a result, items 1-3 or 5-7 shared four features with each other, and six with the target. Vector composition is illustrated in Figure 4. In the relevant experimental conditions the target item was included as item 5 or 3. In the control condition the target was discarded and a new item was generated to be included as item 5 or 3.

Other than the fixed features shared by orthographically similar items we did not assume any differences between the experimental and control conditions, or between the conditions which had items 1-3 or items 5-7 share features. The model therefore contains five free parameters, a, which controls the relationship between proportion of shared features and similarity, λ, which controls how far back retroactive interference operates, r, which controls the effectiveness of rehearsal, s_0, which determines the minimum similarity necessary to trigger a recall, and τ, which controls how deterministically the most similar item is recalled. In addition, the model contains other parameters which are not fitted but are fixed to the values suggested by Neath & Surprenant (2007) for the original Feature Model.

Details of the fitting.

The fitting was done via Approximate Bayes Computation (see Turner & Van Zandt, 2012, or Marin et al., 2012, for a review), as for previous applications of the RFM. We used a version of sequential Monte Carlo sampling known as Partial Rejection Control (Sisson et al., 2007) hereafter referred to as ABC-PRC. Full details are given in the appendix and Code to fit the model can be found on the OSF page. We applied the RFM to the data of Experiment 2 because although the pattern of results was similar in Experiment 1 and 3 and the number of participants in each condition was the same, there were more observations per participant in Experiment 2 than in Experiment 3. The RFM was fit to all four conditions (control lists with target in Position 3, experimental lists with target in Position 3, control lists with target in Position 5, experimental lists with target in Position 5) at once, and for each condition we matched the model to the average proportion of correct items and the error gradients for items 3 and 5.

Although this is not the first time the RFM has been quantitatively fit to data, it is the first instance of the model being fit to error gradient data. This poses something of a challenge for the model and for the fitting procedure. It is possible that despite past success at fitting serial positions, the model might not match error gradients. For the fit, the frequency with which an item presented in, say, Position 3 is recalled in, say, Position 5 is roughly an order of magnitude smaller than the frequency with which the item is recalled in its correct position (see Figure 5) Some experimentation was therefore required to find a weighting of the different data types which produced a suitable discrepancy function.

Results of the fitting are shown in Figure 5. Fits are generally very good and capture the key effects well. It is gratifying to see that the RFM is able to capture error gradients, serial position curves and omission rates with a reasonable degree of accuracy, remembering that only the error gradients and average accuracy rates were fit. Importantly, the RFM correctly predicts the migration of item 3 or 5 towards the group of orthographically similar items in the relevant experimental conditions, and it does this entirely via the similarity between items due to their shared features.

General Discussion

In the introduction to this paper, we briefly reviewed a group of models that have been increasingly influential (see LaRocque et al., 2014). These views insist on the importance of long-term knowledge in producing the behavior that is typically analyzed when studying STM. These ideas led Poirier et al. (2015) to make a controversial suggestion: Order recall errors in STM could be the result of activation perturbations within established lexico-semantic networks.

Although they presented results that supported this idea, there have been critical assessments of the proposal and there is an alternative explanation of the migration results they reported. The latter highlights the features that are shared between semantically related items as well as the features shared by orthographic neighbors.

Here, in order to first assess the generality of the migration effect that Poirier et al. (2015) reported we asked if the migration effect could be produced with another long-term memory factor, namely orthographic neighborhood. Moreover, we ran a study that allowed us to critically contrast the two hypotheses delineated above. Finally, Experiment 3 asked if the migration effects observed here were attributable to phonemic similarity or to orthographic neighborhood as such.

Within experiments 1 and 2, when the target item was in Position 5, we obtained findings that were well aligned with the predictions of ANet, in that the predicted migrations were observed. However, the findings of Experiments 2 and 3 were particularly informative and clearly contradicted the predictions of ANet. Based on ANet, the expectation was that if a target word was *followed* by a number of its neighbors, the increase in the activation produced by the latter items would lead the target word to be more often recalled in an earlier position, relative to corresponding control items.

However, a similarity-based account would predict a different pattern. Specifically, the prediction would be that the target word will migrate more often towards the items that are similar to it, i.e. the neighbors that followed in the list. In other words, in Experiment 2 and 3, ANet and a similarity-based account predict migrations in opposite directions when the target precedes its neighbors. The findings of Experiment 2 and 3 were unequivocal in supporting a similarity-based view. As such, current results are in line with recent work by Kowialiewski et al. (2021) who tested a connectionist model of short-term memory for order that included co-activation and a primacy gradient as in ANet. They concluded that an activation-based order mechanism could not account for the interaction between LTM factors and short-term order memory. The results of Experiment 2 and 3 are also at odds with the order-as-activation idea and instead support the predictions of a similarity-based account (see also Kowialiewski et al., 2021b). Importantly, such an account can also easily explain the results of the first experiment.

To further test the similarity-based account of our findings, we called upon a computational model known as the RFM which has similarity-based retrieval at its heart (Nairne, 1990; Neath & Nairne, 1995; Neath & Surprenant, 2007). The RFM has been used to account for empirical phenomena, both in the verbal and visuo-spatial domain (Poirier et al, 2019). More recently, the RFM was also used to account for the production effect (better recall of items read aloud than silently) in both short- and long-term memory tasks (Cyr et al., 2021; Saint-Aubin et al., 2021). In addition, the model can account for complex interaction patterns between serial positions and list composition in immediate serial recall and delayed free recall, as well as immediate and delayed order reconstruction tasks. Here, we called upon the RFM to model the data obtained in our critical second experiment; the model was made to predict the error gradients related to specific target item migrations as well as accuracy. Overall, the fits were very good; it is also important to note that said fits were made simultaneously across very different scales and measures, given we looked at serial positions curves as well as item transpositions. These findings suggest that a model where relative similarity determines retrieval probability can handle all the varied and detailed data patterns we have reported rather well, as well as account for data from a number of other paradigms.

At the core of the RFM there is a retrieval mechanism whereby a degraded cue is used to attempt retrieval of a response from LTM. The cue is a record of the item, that has been subject to interference by other items and by activity intervening between encoding and retrieval. We note in passing that in the RFM this is identified as a primary memory (PM) representation – However, the ‘location’ of this representation is arbitrarily attributed to PM or STM. Said representation could also be constructed from feature combinations within episodic memory or LTM, as suggested recently by Cowan (2019). Based on this degraded representation, relevant items in LTM compete for retrieval, based on their similarity to the said cue. Any winner of this competition is identified based on relative similarity – that is, absolute similarity or match between the cue and retrieved trace is not the critical factor in successful retrieval. What matters is the relative match. Therefore, if item ‘A’ passes a minimal threshold match and happens to have more similarity to the cue than the other candidates, then item ‘A’ is more likely to be retrieved, even if its absolute level of match is low. Likewise, item ‘B’ could have a high level of similarity to the cue and not be retrieved if most of the other candidates have a comparable level of match to said cue.

The RFM also includes a representation of the order of the items, separate from the representation of the items themselves. It is a mechanism that encodes the order of the items as they are presented with some noise or fuzziness. The mechanism essentially replicates the Estes (1972) Perturbation model of STM order representation, where a single parameter controls the reliability of the order representation or confusability of the positional representations. In this model, also similarity-based, items are much more likely to ‘perturb’, or move, to adjacent positions than to more distal ones. Generally speaking, the model has been very successful in describing the detailed patterns of transposition error frequencies that are typically observed in ordered recall tasks. In a serial recall task, the RFM, at the start of the retrieval process for an item, selects the next item that is output by this perturbation process as the cue for the current retrieval attempt. Importantly however, what the similarity-based retrieval mechanism discussed above implies is that further order errors can easily be produced based on the relative similarity of a target item to other items within the list.

The implication is that for serial order tasks, there is a basic order representation mechanism, but other processes can generate order errors–for instance when an item is recalled in the wrong position because it is selected by the similarity-based retrieval mechanism. Both sources of order errors are similarity-based – or depend on distinctiveness. In the case of order perturbation, the similarity dimension is position within a linear sequence. In the case of later order errors, they are generated because of relative similarity in other task-relevant features.

The above is an important point that illustrates that an explicit model can make clear that observations such as order errors, which are usually assumed to accrue from a single mechanism, can in fact originate from more than one source (see also Neath, 1999).

One of the properties of the proposed mechanism for order memory is that it is reasonably general purpose, in the sense that in the RFM, order representation is the same for verbal, visual, and spatial items (e.g. Poirier et al., 2019). Other models in the field have suggested general mechanism for order encoding and maintenance (Hurlstone & Hitch, 2018), but this is not always the case.

While the account provided by the RFM for order errors is clear, the capacity of a similarity-based model to account for better recall of the target item when it is more similar to other list items remains to be explained. In effect, in the experimental condition, by being an orthographic neighbor of three list items (1-3 or 5-7), the target was more similar, but nevertheless it was better recalled. According to the RFM, this is due to the retroactive interference process in primary memory. Because the target item was an orthographic neighbor of the three items in the experimental condition, 6 of its 20 modality independent features were made to match the corresponding ones in its three neighbors. In our lists, the target item was never followed by an orthographic neighbor: the critical item 3 was followed by an unrelated item in position 4 and the critical item 5 was followed by an unrelated item in position 6. Therefore, the modality independent features representing the orthographic neighborhood would be relatively immune to retroactive interference. With a better cue in primary memory, the probability of sampling the appropriate representation in secondary memory would be higher. It is worth noting that when the target item was in Position 5, the orthographic neighbor in Position 3 was not better recalled than the corresponding item in the control condition. This might be expected, as that item is not followed by another similar neighbor. Hence, relative to the neighbors in positions 1 and 2, this 3rd neighbor will not have as many features overwritten by the subsequent fourth item. However, we did not have a control condition with the first four items within the list being non-neighbors. The control condition also has three neighbors of an item that is not presented in Position 5. Therefore, item 3 is recalled to a comparable level in both the experimental and control conditions. This is predicted by the model, because in both the experimental and the control condition, the first three items are orthographic neighbors. Therefore, the last item of this group, in both conditions, enjoys the same benefit of not being followed by another neighbor.

Taken together the findings we reported here allow us to discard the ANet model in favor of the RFM. ANet was based on the idea that a primacy gradient, often invoked to explain the shape of the serial position curve in immediate serial recall (see, e.g., Hurlstone et al., 2014), was the results of activation within a lexico-semantic network in LTM. The results of Experiments 2 and 3 in particular are not compatible with the predictions of ANet. This does not imply that the proposed primacy gradient is not an important part of order representation; indeed, the order retrieval mechanism included in the RFM has an implicit primacy gradient as the first item is selected first, unless it has perturbed, followed by the second item, etc. However, the proposal that this primacy gradient is the result of activation and co-activation between the items was not supported by the findings reported here (see also, Kowialiewski et al., 2021).

Instead, the RFM, a model that heavily relies on retroactive interference and relative similarity to account for retrieval, was much more successful. As mentioned earlier, the model has also been applied to a series of other tasks and timeframes (delayed reconstruction, free recall, visuo-spatial STM) and hence has some generality (Cyr et al., 2021; Poirier et al., 2019; Saint-Aubin et al., 2021). In terms of the implications for order representation, the model assumes each item is encoded with features that indicate its position, but that this information can perturb or change, over time. These assumptions, along with the simple quantitative uncertainty formulae from the Estes' (1972) proposal produce realistic order recall behavior, both in the current study and a number of others (Nairne, 1991; Neath, 1999, 2000). We note in passing that in their current instantiations, neither the Primacy model (Page & Norris, 1998; Norris, Kalm, & Hall, 2019) nor the Start-End model (Henson, 1998) can account for most of the findings reported here, although the RFM achieved good fits.

Conclusion

Previous interpretations have insisted that long-term memory factors have almost all of their effect by increasing item recall irrespective of position (Saint-Aubin & Poirier, 1999a, 1999b). However, a number of studies have reported statistically reliable effect of long-term memory factors on order errors (see, e.g., Hulme et al., 2003; Saint-Aubin et al., 2005; Saint-Aubin & Poirier, 2000; Tse & Altarriba, 2007). The RFM implemented here offers a straightforward and parsimonious interpretation of this typical pattern of findings and highlights that order errors can have multiple origins. In the RFM, the representation of the words in an immediate serial recall task relies on a basic order representation mechanism where the positional encoding of items has a certain probability of drifting in either direction, with the drift increasing as time –and hence opportunity to perturb—increases. Importantly however, if item ‘X’ is quite similar to item ‘Y’, it can mistakenly be recalled in item Y’s stead – producing an order error through another route. The latter easily explains why items that share phonological or semantic features, for example, will tend to generate more order errors than items that are less similar. Perhaps more importantly, the successful application of the RFM to multiple tasks, timeframes and both verbal and visuo-spatial materials highlights the usefulness of the model and the power of the relative distinctiveness mechanism that is the main engine of retrieval within.

We started off with the predictions of an activation-based view which produced a counterintuitive prediction i.e., ANet. The latter was not supported by the data, however. In order to account for results, a different kind of model was called upon, one where activation is not required - except perhaps in terms of selecting recall candidates - and where the to-be-explained effects appear because of a retrieval mechanism that is based on relative distinctiveness. Overall, this work has highlighted a new item migration phenomenon and further underlines the strong and complex interactions between long-term knowledge and memory over the short-term.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*, 50-68.

Acheson, D.J., MacDonald, M.C. & Postle, B.R. (2011). The effect of concurrent semantic categorization on delayed serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 44-59.

Baddeley, A. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Bhatarah, P., Ward, G., Smith, J., Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition, 37*, 689-713. https://doi.org/10.3758/MC.37.5.689

Buchsbaum B.R. & D'Esposito M. 2008. The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience*, *20*, 762-778.

Camos, V. (2015). Storing verbal information in working memory. *Current Directions in Psychological Science, 24,* 440-445. doi: 10.1177/0963721415606630

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge, U.K.: Cambridge University Press.

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin, 145(8),* 822–847. https://doi.org/10.1037/bul0000199

Cowan, N., & Chen, Z. (2009). How chunks form in long-term memory and affect short-term memory limits. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 86-101). Hove, East Sussex, UK: Psychology Press.

Cyr, V., Poirier, M., Yearsley, J. M., Guitard, D., Harrigan, I., & Saint-Aubin, J. (2021). The production effect over the long term: Modeling distinctiveness using serial positions. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0001093.supp (Supplemental)

Derraugh, L. S., Neath, I., Surprenant, A. M., Beaudry, O., & Saint-Aubin, J. (2017). The effect of lexical factors on recall from working memory: Generalizing the neighborhood size effect. *Canadian Journal of Experimental Psychology, 71*, 23-31. https://doi.org/10.1037/cep0000098

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*, 190-193. doi:10.1177/1745691617720478

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds.), *Coding Processes in Human Memory*. Washington, D.C.: Winston.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Guitard, D., Saint-Aubin, J., Tehan, G., & Tolan, A. (2018). Does neighborhood size really cause the word length effect? *Memory & Cognition, 46(2)*, 244–260. https://doi.org/10.3758/s13421-017-0761-9

Guérard, K., & Saint-Aubin, J. (2012). Assessing the effect of lexical variables in backward recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 312-324.

Guest, O. & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691620970585

Gupta, P. (2009). A computational model of non word repetition, immediate serial recall, and non-word learning. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain*. Hove: Psychology Press

Henson, R. N. A. (1998). Short-term memory for serial order: The Start–End Model. *Cognitive Psychology*, *36(2)*, 73–137. https://doi.org/10.1006/cogp.1998.0685

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

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49(4)*, 500–518. https://doi.org/10.1016/S0749-596X(03)00096-2

Hurlstone, M. J., & Hitch, G. J. (2018). How is the serial order of a visual sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44(2)*, 167–192. https://doi.org/10.1037/xlm0000440.supp (Supplemental)

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological*

Bulletin, 140(2), 339–373. https://doi.org/10.1037/a0034221

Kowialiewski, B., Gorin, S., & Majerus, S. (2021b). Semantic knowledge constrains the processing of serial order information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Kowialiewski, B., Lemaire, B., Majerus, S., & Portrat, S. (2021). Can activated long-term memory maintain serial order information? *Psychonomic Bulletin & Review, 28(4),* 1301–1312. https://doi.org/10.3758/s13423-021-01902-3

LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience, 8.*

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: the importance of distinguishing item and order information. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244-276). Hove, UK: Psychology Press.

Majerus, S., Van der Linden, M., Mulder, L., Meulemans, T., Peters, F. (2004) Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *51*, 297-306

Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing, 22(6),* 1167–1180. https://doi.org/10.1007/s11222-011-9288-2

Martin, N. & Gupta, P. (2004). Exploring the relationship between word processing and verbal STM: Evidence from associations and dissociations. *Cognitive Neuropsychology*, *21*, 213-228.

Martin R. C. (2006). Semantic short-term memory, language processing, and inhibition. In Meyer A. S., Wheeldon L. R., Krott A. (Eds.), *Automaticity and control in language processing* (pp. 162-191). Hoboken, NJ: Taylor & Francis.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*, 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18(3)*, 251–269. https://doi.org/10.3758/BF03213879.

Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition, 19(4)*, 332–340. https://doi.org/10.3758/BF03197136

Neath, I. (1999). Modelling the disruptive effects of irrelevant speech on order information. *International Journal of Psychology, 34*, 410-418. https://doi.org/10.1080/002075999399765

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7(3)*, 403–423. https://doi.org/10.3758/BF03214356

Neath, I., Hockley, W. E., & Ensor, T. M. (2021). Stimulus-based mirror effects revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000901

Neath, I., & Nairne, J. S. (1995). Word-Length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review, 2*, 429–441. https://doi.org/10.3758/BF03210981.

Neath, I., Saint-Aubin, J., & Surprenant, A. M. (2022). Semantic relatedness effects in serial recall but not in serial reconstruction of order. *Experimental Psychology*.

Neath, I., & Surprenant, A. M. (2007). Accounting for age-related differences in working memory using the feature model. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory: Behavioural and neural correlates* (pp. 165–179). Oxford, UK: Oxford University Press.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments & Computers, 36(3)*, 516–524. https://doi.org/10.3758/BF03195598

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin, 143(9)*, 992–1009. https://doi.org/10.1037/bul0000108

Norris, D. (2019). Even an activated long-term memory system still needs a separate short-term store: A reply to Cowan (2019). *Psychological Bulletin, 145(8)*, 848–853. https://doi.org/10.1037/bul0000204

Norris, D., Kalm, K., & Hall, J. (2020). Chunking and redintegration in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(5)*, 872–893. https://doi.org/10.1037/xlm0000762

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144(9)*, 885–958. https://doi.org/10.1037/bul0000153

Saint-Aubin, J., Poirier, M., Yearsley, J. M., Guitard, D., & Robichaud, J.-M. (2022, August 13). Order Recall in Verbal Short-Term Memory: A Walk Around the Neighborhood.

Retrieved from osf.io/ga49s

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105(4)*, 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *48A*, 384-404

Poirier, M. & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology*, *50*, 408-412.

Poirier, M, Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43*, 489-499.

Poirier, M., Yearsley, J. M., Saint-Aubin, J., Fortin, C., Gallant, G., & Guitard, D. (2019). Dissociating visuo-spatial and verbal working memory: It's all in the features. *Memory & Cognition, 47(4)*, 603–618. https://doi.org/10.3758/s13421-018-0882-9

Roodenrys, S. (2009). Explaining phonological neighbourhood effects in short-term memory. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 177-197). Hove: Psychology Press

Roodenrys, S., Guitard, D., Miller, L. M., Saint-Aubin, J., & Barron, J. M. (2022). Phonological similarity in the serial recall task hinders item recall, not just order. *British Journal of Psychology*. https://doi.org/10.1111/bjop.12575

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1019–1034.

Roodenrys, S., Miller, L. M., & Josifovski, N. (2022). Phonemic interference in short-term memory contributes to forgetting but is not due to overwriting. *Journal of Memory and Language*, 122. https://doi.org/10.1016/j.jml.2021.104301

Saint-Aubin, J., Ouellette, D. & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials? *Psychonomic Bulletin & Review, 12*, 171-177.

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

Saint-Aubin, J. & Poirier, M. (1999b). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology, 52A*, 367-394

Saint-Aubin, J. & Poirier, M. (2000). Immediate serial recall of words and non-words: a test of the retrieval based hypothesis. *Psychonomic Bulletin & Review*, *7*, 332-340

Saint-Aubin, J., Poirier, M., Yearsley, J. M., Guitard, D., & Robichaud, J.-M. (2022, August 13). Order Recall in Verbal Short-Term Memory: A Walk Around the Neighborhood. Retrieved from osf.io/ga49s

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2021.104219

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94(2)*, 143–174. https://doi.org/10.1348/000712603321661859

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*, 76-80.

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America, 104(6),* 1760–1765. https://doi.org/10.1073/pnas.0607208104

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2009). Correction for Sisson et al., sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America, 106(39),* Article 16889. https://doi.org/10.1073/pnas.0908847106

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*, 1096 –1104. http://dx.doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolKit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44,* 24 –31. http://dx.doi.org/10.1177/0098628316677643

Thorn A. S. C., Frankish C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 729-735.

Tse, C.-S., Altarriba, J. (2007). Testing the associative-link hypothesis in immediate serial recall: Evidence from word frequency and word imageability effects. *Memory, 15,* 675-690.

Turner, B. M., & Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology, 56(2),* 69–85. https://doi.org/10.1016/j.jmp.2012.02.005

Table 1

Samples of the experimental and the control lists for Experiments 1, 2 and 3.

			Experiment 1			
Experimental List Examples						
plaie	**glace**	**plage**	chambre	**place**	guerre	pièce
[wound]	**[ice]**	**[beach]**	[bedroom]	**[location]**	[war]	[room]
coude	**poule**	**colle**	druide	**coule**	flamme	plaque
[elwbow]	**[hen]**	**[glue]**	[druid]	**[flow]**	[flame]	[plate]
Control List Examples						
paix	main	paon	saut	soin	chair	thé
[peace]	[hand]	[peacock]	[jump]	[care]	[flesh]	[tea]
soif	foin	soir	haut	pain	tract	vol
[thirst]	[hay]	[evening]	[high]	[bread]	[leaflet]	[flight]
			Experiment 2			
Experimental List Examples						
chambre	guerre	**place**	pièce	**plaie**	**glace**	**plage**
[bedroom]	[war]	**[location]**	[room]	**[wound]**	**[ice]**	**[beach]**
coude	**poule**	**colle**	druide	**coule**	flamme	plaque
[elwbow]	**[hen]**	**[glue]**	[druid]	**[flow]**	[flame]	[plate]
Control List Examples						
saut	chair	soin	thé	paix	main	paon
[jump]	[flesh]	[care]	[tea]	[peace]	[hand]	[peacock]
soif	foin	soir	haut	pain	tract	vol
[thirst]	[hay]	[evening]	[high]	[bread]	[leaflet]	[flight]
			Experiment 3			
Experimental List Examples						
kit	**mit**	**art**	gus	**ait**	nom	duo
/kit/	/mi/	/aʁ/	/gys/	/ɛ/	/nɔ̃/	/dɥo/
[set]	[put]	[art]	[guys]	[have]	[name]	[duet]
cou	ému	**oit**	ban	**dit**	**ont**	**lit**
/ku/	/emy/	/wa/	/bɑ̃/	/di/	/ɔ˜/	/li/
[neck]	[moved]	[hears]	[banns]	[says]	[have]	[bed]
Control List Examples						
hum	hie	hui	sac	ait	zoo	ove
/ œm/	/i/	/ɥi/	/sak/	/ɛ/	/zɔo/	/ɔv/
[hem]	[rammer]	[boo]	[bag]	[have]	[zoo]	[ovum]
mou	tub	dot	sec	oie	vie	âge
/mu/	/tœb/	/dɔt/	/ sɛk/	/wa/	/vi/	/ aʒ/
[soft]	[tub]	[dowry]	[dry]	[goose]	[life]	[age]

Note. In Experiment 1, in the experimental condition, the words in bold represent the orthographic neighbors (position 1, 2, and 3) and the target word (Position 5). In Experiments 2 and 3, in the experimental condition, the words in bold represent the orthographic neighbors (position 5, 6, and 7 or position 1, 2, and 3) and the target word (Position 3 or Position 5). In Experiment 3, phonetic transcriptions are provided under the French words.

Table 2

Mean proportion and standard deviation in parentheses of correct recall with a strict and a free recall criterion, for the critical item (3 or 5) and for all other list items as a function of condition (control vs. experimental) and experiment.

	Correct in position scores		Item recall scores	
	All other positions	Critical item	All other positions	Critical item
Condition	*M* (*SD*)	*M* (*SD*)	*M*(*SD*)	*M* (*SD*)
Experiment 1				
Critical item 5				
Control	.43 (.13)	.16 (.16)	.49 (.12)	.27 (.18)
Experimental	.41 (.12)	.19 (.18)	.48 (.12)	.39 (.17)
Experiment 2				
Critical item 3				
Control	.39 (.17)	.54 (.25)	.53 (.14)	.62 (.22)
Experimental	.40 (.16)	.57 (.26)	.53 (.13)	.69 (.21)
Critical item 5				
Control	.43 (.17)	.19 (.21)	.56 (.13)	.37 (.24)
Experimental	.44 (.18)	.21 (.23)	.57 (.13)	.48 (.23)
Experiment 3				
Critical item 3				
Control	.40 (.17)	.30 (.24)	.57 (.13)	.38 (.25)
Experimental	.41 (.19)	.37 (.27)	.60 (.15)	.48 (.25)
Critical item 5				
Control	.48 (.18)	.18 (.20)	.62 (.14)	.28 (.23)
Experimental	.50 (.18)	.24 (.22)	.64 (.13)	.37 (.23)

Figure 1

Proportion of trials showing an error for Item 5 as a function of presentation position in Experiment 1; only the erroneous recall positions are plotted on the x-axis. Error bars represent 95% within-participant confidence intervals computed according to Morey's (2008) procedure.

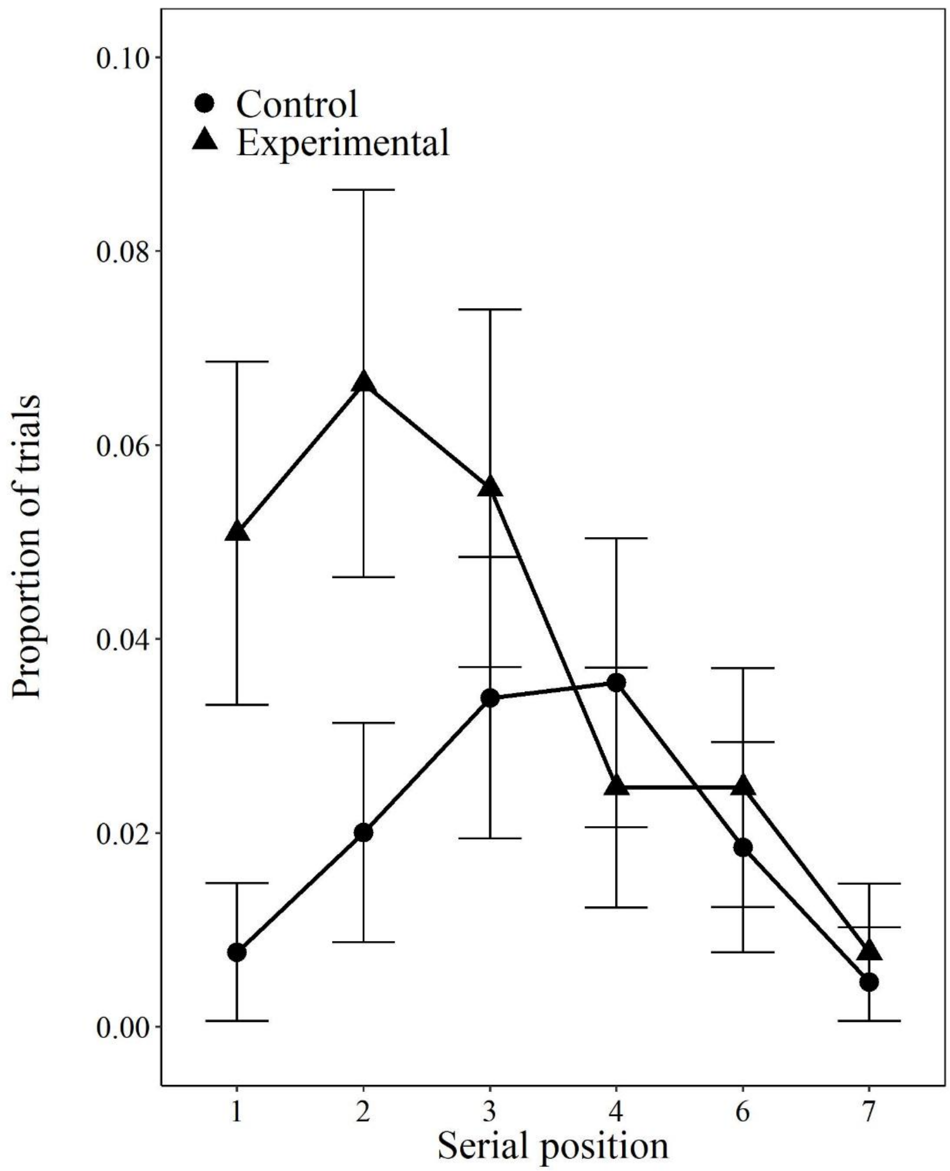


Figure 2

Proportion of trials showing an error for Item 3 (left panel) and Item 5 (right panel) as a function of presentation position in Experiment 2; only the erroneous recall positions are plotted on the x-axis. Error bars represent 95% within-participant confidence intervals computed according to Morey's (2008) procedure.

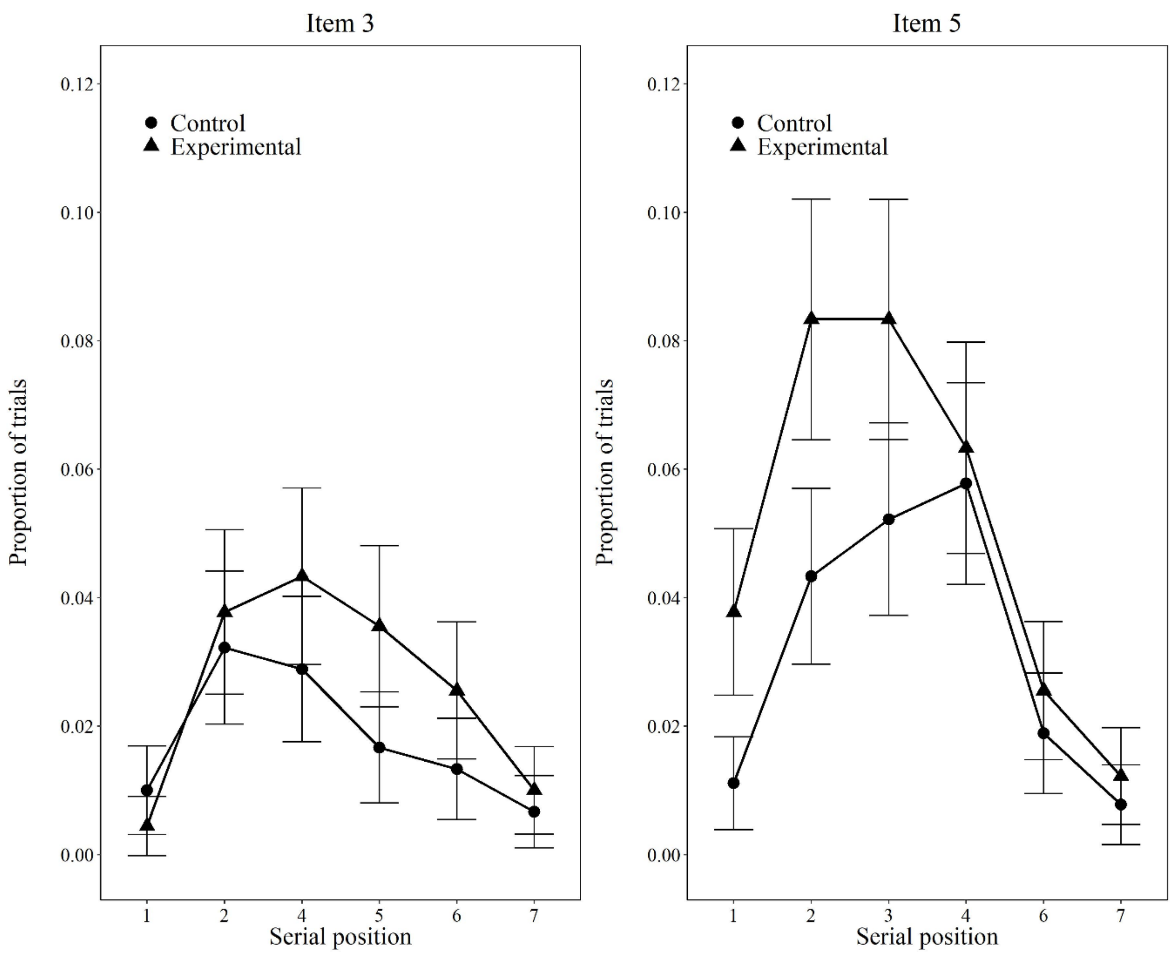


Figure 3

Proportion of trials showing an error for Item 3 (left panel) and Item 5 (right panel) as a function of presentation position in Experiment 3; only the erroneous recall positions are plotted on the x-axis. Error bars represent 95% within-participant confidence intervals computed according to Morey's (2008) procedure.

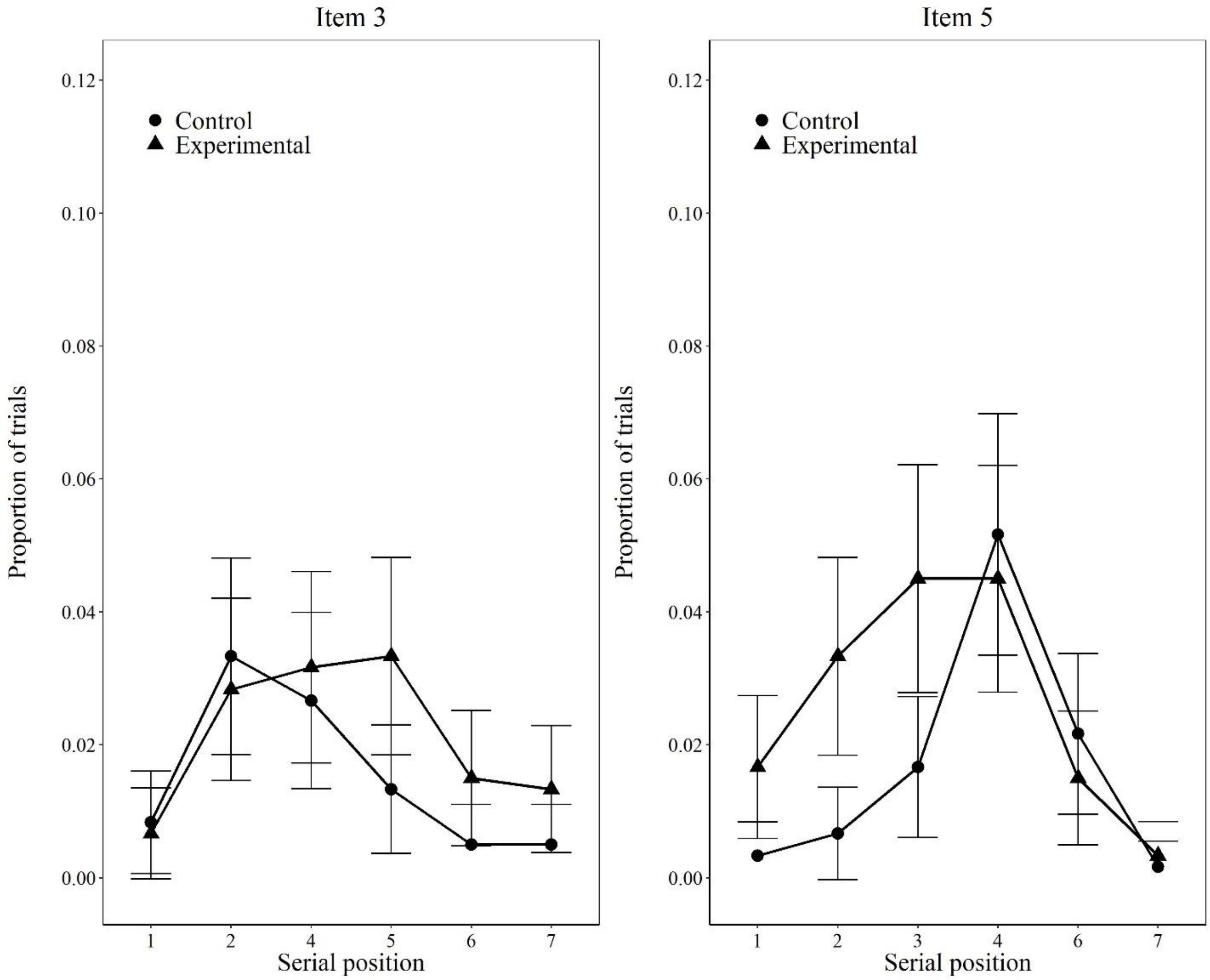


Figure 4

Illustration of vector composition in the Revised Feature Model. Each multi-colored column represents an item vector in which colored rectangles stand for distinct features. For illustrative purposes, items contain fewer features than used in the model. The yellow, green and pink rectangles represent the values of 1, 2, and 3. In the experimental condition, the broken rectangle represents the shared features of the three orthographic neighbors with the target, and in the control condition the rectangle represents the shared features of the three orthographic neighbors of a non-presented target. For each item, the dashed rectangle represents a feature which is not shared with the target (presented in the experimental condition and absent from the control condition). The numbers at the bottom represent the position of the item in the list.

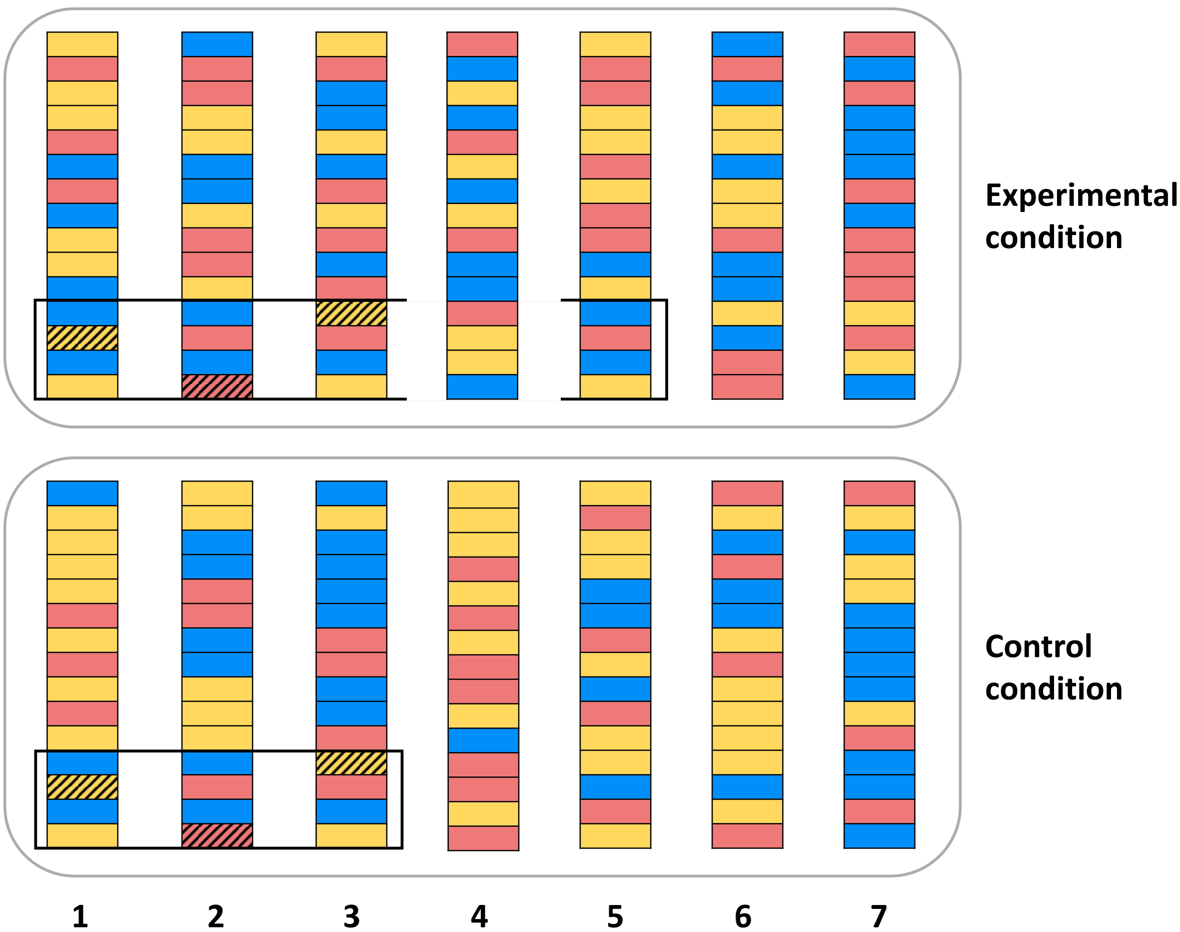
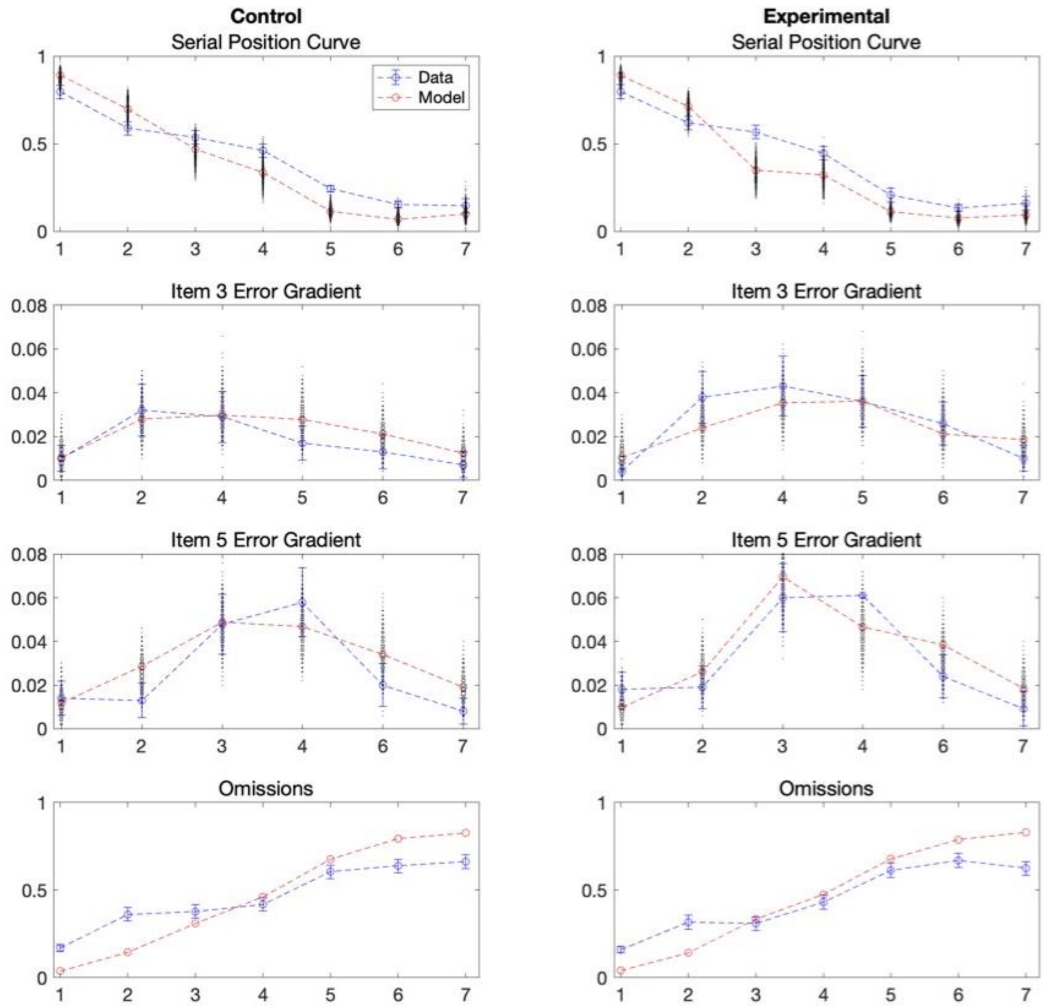
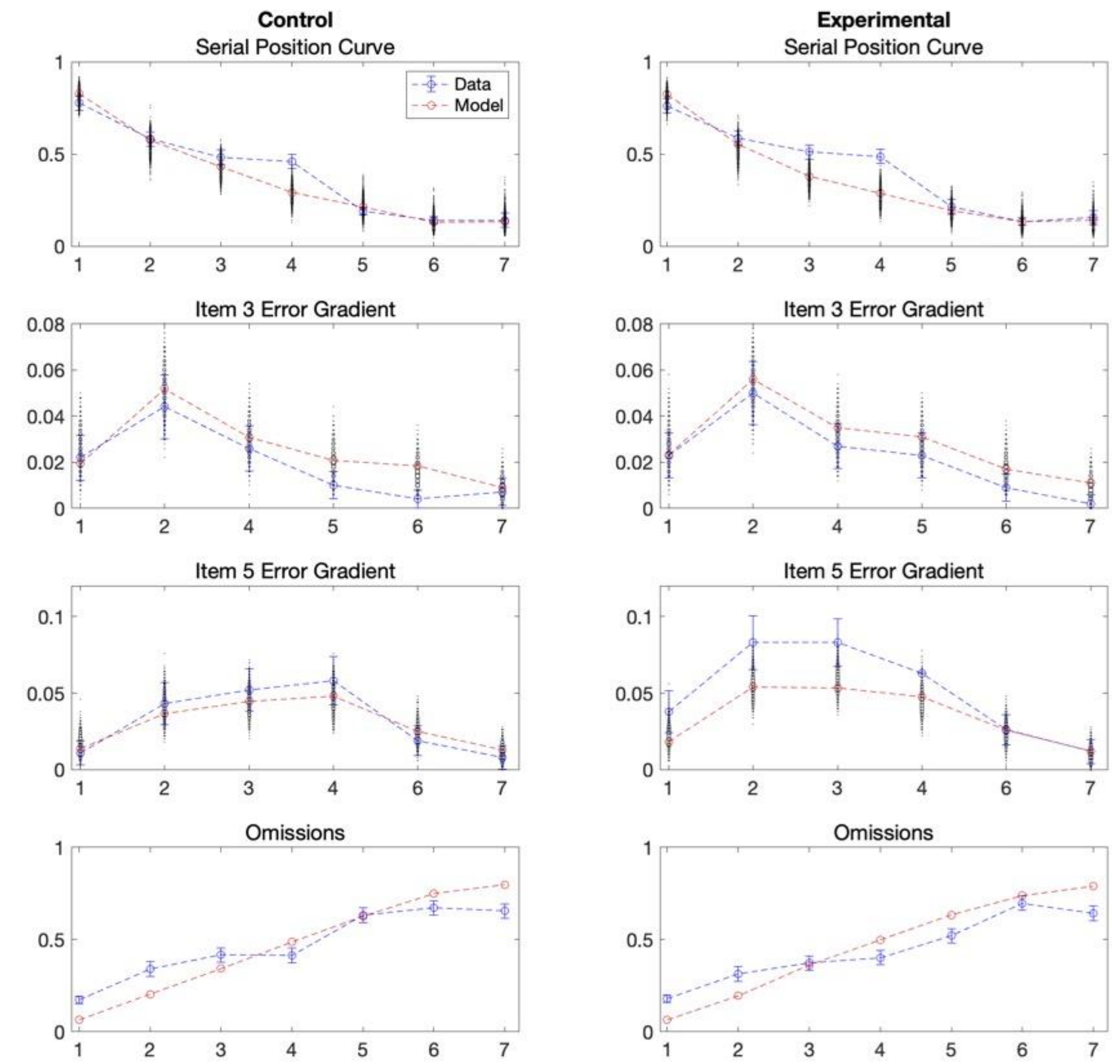


Figure 5

Data (Blue line) vs Model best fit (Red line) and Model Fitting posterior (Black points) for Experiment 2.



Panel a: *Item 3 condition. Left column shows Serial Position curve, Error Gradients for Items 3 and 5, and Omissions for the Control condition. Right column shows the same data for the experimental condition. Fits are generally very good and capture all features of the data well.*



Panel b: *Item 5 condition. Left column shows Serial Position curve, Error Gradients for Items 3 and 5, and Omissions for the Control condition. Right column shows the same data for the experimental condition. Fits are generally good and capture the data well, although there is a somewhat smaller effect in the Item 5 error gradient than expected.*

Appendix: Model Fitting Details.

The RFM is too complex for an analytic expression for the likelihood to be derived, so as with all previous attempts to fit the model to data we used a version of Approximate Bayesian Computation (ABC) (see Turner & Van Zandt, 2012, or Marin et al., 2012, for a review). Following Poirier et al. (2019), Saint-Aubin et al. (2021), and Cyr et al (2021) we used ABC Partial Rejection Control (ABC-PRC) (Sisson et al., 2007, 2009). ABC-PRC works by repeatedly sampling from a prior over the parameter space until it finds a set of parameters which generate a set of summary statistics (in our case error gradients, and mean accuracy and omission rate) sufficiently close to the data, as determined by the discrepancy function. When this happens, the algorithm stores these parameter values, and moves on to the next particle in the generation. Once all particles in a generation have been associated with parameter sets, the algorithm gives each particle a weight depending on the prior, and then begins a new generation, sampling from the previous generation with probabilities given by the weights, and repeatedly perturbing around the previous parameter values until a set is found producing summary statistics even closer to the data. Once the required number of generations have elapsed posterior estimates for the parameters can be obtained as the fraction of particles in the final generation with that parameter value. Posterior predicted distributions of the summary statistics are also easily obtained. For full details see Sisson et al. (2007) (Note also the errata, Sisson et al., 2009).

The choice of discrepancy function is not usually an issue for much discussion in ABC, a simple sum-squared error is often adequate. However, things are made more complex here by the fact we have to fit to mean accuracy and error gradients together. The reason this is challenging is it requires us to specify a distance function across these two different data spaces. Looking at the data from Experiment 2, typical Item errors are roughly 2% and typical points on the serial position curve are around 50%. We might therefore consider simulated item error of 2.2% to be as distant from the true value of 2% as a simulated accuracy of 55% is from the true accuracy of 50%. This gives a relative weighting of (0.5/0.02)^2=625. However, the average accuracy is computed over 7 positions, so there is an extra factor of 7^2, and meanwhile there are 12 item errors, so the weighting picks up an additional factor of ~4, giving an overall weighting of around 2000, which was the number used in fitting. Note that all this factor does in practice is control where to prioritize the fit between model and data, in the serial position curve or the item error gradients.

The important parameters for ABC-PRC are the number of particles (set to 1000 for all fits reported here), the details of the prior, the proposal distributions, and the minimum tolerances for each fit. The proposal and tolerances can be found in the code on the OSF. Priors, and resulting posterior distributions are summarized in Table A1.

Parameter	Prior	Best Fit Value (Median, 95% HDI)
Overwriting parameter, λ	$Normal(1,0.3)$	1.15 [0.65, 1.82]
Distance scaling parameter, a	$Normal(3,1)$	3.91 [2.65, 5.51]
Rehearsal Parameter, r	$Beta(1.5,1.5)$	0.84 [0.60, 0.99]
Temperature parameter, τ	$HalfNormal(0,0.3)$	0.15 [0.09, 0.20]
Minimum Similarity, S_0	$HalfNormal(0,1)$	0.27 [0.14, 0.39]

Table A1: Median and 95% HDI of the parameter posteriors for fits to Experiment 3.

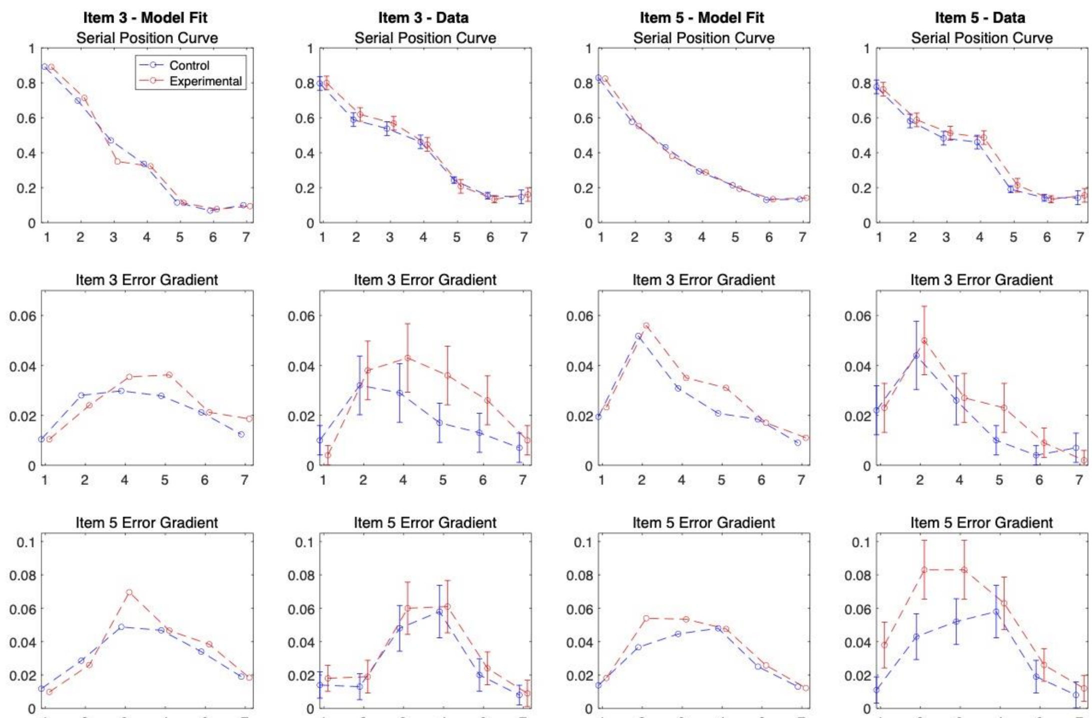


Figure A1: Comparison of Data and Model Fits for each condition. Displaying data and model fits in this way helps us understand the extent to which the model matches the qualitative features seen in the data. The key panels are the Item 3 error gradient for the Item 3 condition, and the Item 5 error gradient for the Item 5 condition. The data in each case illustrates the key property that in the experimental condition the target item (eg item 3 in the Item 3 condition) is more likely to be recalled in place of one of its orthographic neighbors (eg items 5-7 in the Item 3 condition). We can see that this qualitative behavior is reproduced by the model, although the size of the effect is somewhat smaller.