The Interaction Between the Production Effect and Serial Position in Recognition and Recall

Sébastien Gionet

Université de Moncton

Dominic Guitard

Cardiff University

Jean Saint-Aubin

Université de Moncton

Authors' Note

This research was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to JSA. While working on the manuscript, SG was supported by a doctoral scholarship from NSERC. Correspondence concerning this article should be addressed to Sébastien Gionet at esg8807@umoncton.ca, Dominic Guitard at guitardd@cardiff.ac.uk or Jean Saint-Aubin at jean.saint-aubin@umoncton.ca

Abstract

In memory tasks, items read aloud are better remembered than their silently read counterparts. This production effect is often interpreted by assuming a distinctiveness benefit for produced items, but whether this benefit also comes at a cost remains up for debate. In recall tasks, when pure lists are used in which all items are produced or read silently, studies have shown a better recall of produced items at the last serial positions, but a lower recall at the first positions. This cost of production has been interpreted by assuming that production interferes with rehearsal. However, in recognition tasks, models typically assume that the distinctiveness benefit for produced items comes at no cost. Across four experiments, participants completed a 2AFC recognition test, an old-new recognition test or an immediate serial recall test. List length was also manipulated. Results show that although the production effect is larger at the last serial positions, the cross-over interaction between the production effect and serial position observed in recall was not present in recognition. These results suggest that task-related differences in the production effect may inform us about the modulation of basic memory processes by task demands.

Keywords. Production Effect, Serial Positions, Short-term Memory, Item Recognition, Immediate Serial Recall

The Interaction Between the Production Effect and Serial Position in Recognition and Recall

Enhancing the capacity of our own memory is a universal aspiration. Unfortunately, most effective means for achieving this goal are both time and effort consuming (e.g., the pegword method, in which to-be-remembered words are linked to a pre-learned rhyme of words and numbers), thus compromising their widespread usage. In the laboratory, a simple strategy has been found to have a large beneficial effect on memory. That is, reading words aloud at encoding improves memory compared to reading them silently (for an overview see, MacLeod & Bodner, 2017). This production effect has been observed across a variety of tasks including free recall (e.g., Schwell et al., 2024; Lambert et al., 2016), order reconstruction (e.g., Jonker et al., 2014; Saint-Aubin et al., 2021), item recognition (e.g., Hourihan & Fawcett, this issue; MacLeod et al., 2010), and immediate serial recall (e.g., Saint-Aubin et al., 2021). In addition, when pure lists are used (i.e., lists in which all words are read aloud or silently), the production effect has been found to interact with serial position. More specifically, as shown in Figure 1, in order reconstruction, immediate serial recall and delayed free recall, produced items are better recalled than silently read items on the recency portion of the serial position curve, but less well recalled on the primacy portion (e.g., Dauphinee et al., 2024; Gionet et al., 2022; Grenfell-Essam et al., 2017; Macken et al., 2016; Murray et al., 1974; Saint-Aubin et al., 2021).

This interaction between the production effect and serial position has been interpreted as reflecting the costs and benefits of increasing the relative distinctiveness of produced items (Cyr et al., 2022; Dauphinee et al., 2024; Gionet et al., 2022; Saint-Aubin et al., 2021). Specifically, within the Revised Feature Model (RFM; Saint-Aubin et al., 2021), reading words aloud would hinder their rehearsal because explicit articulation and rehearsal share the same resources (e.g., Murray, 1967). This would then selectively impact the retrieval of the first items, which usually benefit more from rehearsal than the last ones (Bhatarah et al., 2009; Rundus, 1971; Tan & Ward, 2008). However, compared to silent reading, reading aloud would also generate extra features through articulation and auditory feedback (see, MacLeod et al., 2010, 2022). These additional features would increase the relative distinctiveness of produced items, and since the RFM assigns forgetting to retroactive interference, the last produced items in a pure list would benefit more from them (Cyr et al., 2022; Dauphinee et al., 2024; Saint-Aubin et al., 2021).

As of yet, the RFM has been applied to immediate serial recall, order reconstruction, free recall, and immediate spatial recall (see, e.g., Cyr et al., 2022; Saint-Aubin et al., 2021, 2024). In verbal recall, as shown in Figure 1, the interaction with serial position is ubiquitous. Furthermore, two systematic reviews revealed that the production effect systematically interacts with serial position in free recall and immediate serial recall (Dauphinee et al., 2024; Gionet et al., 2022). In addition, Dauphinee et al. tested the hypothesis that the cost of producing the items derives from hindered rehearsal. They did so by manipulating presentation speed, because it is well-established that rehearsal is less efficient with a rapid presentation (Bhatarah et al., 2009; Grenfell-Essam & Ward, 2012). As predicted, the disadvantage for produced items at the first serial positions was observed with slow presentation rates allowing the rehearsal of silent items, but not with the fast presentation rate preventing the rehearsal of those items.

Importantly, the view that producing items increases their relative distinctiveness but comes at a cost is not commonly found (see, Saint-Aubin et al., 2021). In fact, most models accounting for the production effect have only been applied to item recognition tasks (Caplan & Guitard, 2024; Jamieson et al., 2016; Kelly et al., 2022; Wakeham-Lewis et al., 2022). Moreover, performance in item recognition is not typically assessed as a function of serial position, and studies of the production effect are no exception (see, Fawcett et al., 2023). In those studies, a small advantage for produced items in pure lists is usually found, but with no noticeable cost

(Fawcett, 2013). Accordingly, Jamieson et al. (2016) modeled the production effect using MINERVA 2 by assuming that produced items benefit from five more features than silently read items. Those additional features were thought to be related to sensory feedback. The model also assumes that retrieval is cue-driven and that each trace is activated in proportion to its similarity to the retrieval cue. In this context, the additional features would increase the distinctiveness of the produced items' memory trace. However, to the best of our knowledge, MINERVA 2 assumes that those features would come at no cost. Similarly, Kelly et al. (2022) and Wakeham-Lewis et al. (2022) used the Retrieving Effectively from Memory framework (REM; Shiffrin & Steyvers, 1997), a computational model of recognition memory, to account for the production effect. Both groups assumed that produced items benefit from ten more encoded features that can be used at retrieval, but once more, the REM does not include any cost from the addition of those features.

In this issue, Caplan and Guitard account for the production effect using the attentional subsetting theory, which posits that only a small subset of item features are attended in episodic recognition tasks. This theory distinguishes between early-emerging shallow features such as orthographic and phonologic attributes in a densely packed feature space, and deep features like semantic attributes that emerge later in a sparser space. It posits that production specifically influences the encoding of some features (e.g., phonologic features in spoken words), which increases the vector length, thereby enhancing the overall strength and effectiveness of memory encoding. By increasing the number of attended features, production boosts the distinctiveness of produced items resulting in a memory advantage for those items. Caplan and Guitard also explore the potential for a trade-off in memory encoding, echoing concepts from the RFM (Saint-Aubin et al., 2021). Specifically, production might divert attention from orthographic and deeper semantic features, potentially leading to fewer of these features being encoded. This trade-off could also be theorized to be consistent across serial positions, but is not currently implemented.

As could be seen, there is an important difference between recognition models which are either silent about a cost of production or suggesting no interaction with serial position (Caplan & Guitard, 2024; Kelly et al., 2022; Jamieson et al., 2016; Wakeham-Lewis et al., 2022) and the only account in recall suggesting an interaction (Saint-Aubin et al., 2021). The interaction between the production effect and serial position in item recognition was recently investigated by Wakeham-Lewis et al. (2022). In their study, they assessed the production effect across all the 90 possible serial positions and found a trend for a larger production effect toward the end of the list. However, they failed to find credible evidence for an interaction or an effect of serial position. Subsequently, Fawcett et al. (2023) addressed the same issue in their meta-analysis. Because list length varied across the nine available datasets, they reported mean recognition performance for the initial three items (early) and the final three items (late). Just as it had been observed with recall tasks, they found a large recency advantage for produced items. However, they also found a smaller advantage for produced items at the first positions. In light of their findings, they stated that "presently, we can conclude only that the between-subjects production effect in recognition appears to be marginally more pronounced for late items relative to early items." (p.42).

Taken together, results so far suggest an interaction between the production effect and serial position in item recognition (Fawcett et al., 2023), but whether a cost of production can be observed on the first serial positions remains up for debate. It should be noted that in all but one of the studies used by Fawcett et al. (2023) for their meta-analysis, serial positions were not reported, analyzed or even discussed, with Forrin et al. (2019) being the exception. However, in that study, multiple blocks of five items were presented and serial positions were only reported as a function of position within the block rather than position across the list. In Wakeham-Lewis' et al. (2022) study, not included in Fawcett et al.'s (2023) meta-analysis, the interaction with serial position was assessed as a control to rule out a practice effect.

Therefore, the designs of previous studies were never optimized to reveal any cost of production on the first serial positions. However, it has been shown that when two distinct tasks are performed under similar methodologies, performance can be much more similar than what is usually found with larger methodological differences (Grenfell-Essam et al., 2017). In the current study, we offer the first critical test of a deleterious production effect on the early serial positions in item recognition when conditions are selected to maximize the chances of observing such an effect. To this aim, in addition to the typical random presentation of the items at test, we tested items in their presentation order, because it has been shown that differences across memory tasks can be driven by output order (Ensor et al., 2020; Ward et al., 2010). Furthermore, because list length has been found to be a critical factor producing dissociations, in Experiments 2 and 3, we used 6-item lists as in typical immediate serial recall experiments (e.g., Grenfell-Essam et al., 2017). Finally, to investigate any task-related differences and to allow for a direct comparison between recall and recognition with regard to the interaction between the production effect and serial position, we used an immediate serial recall task in Experiment 4.

Experiment 1

Method

Sample Size Calculation and Participants

Our sample size for all experiments was based on an a priori power analysis computed with G*Power 3.1.9.7 (Faul et al., 2007). First, the analysis for a within-subject ANOVA with a medium effect size (f = 0.25; Cohen, 1988, p. 286), one group and 2 measurements revealed that 34 participants would be enough to achieve 80% power at the .05 level of significance (see OSF for the full G*Power analysis output). However, due to uncertainties associated with the change in memory task, we slightly overpowered our design. Therefore, a post hoc power analysis with a sample of 48 participants and the same parameters revealed 92% power to detect a medium-sized interaction effect. We recruited 48 undergraduate students (31 women, 17 men, *M* age = 20.3 years, *SD* = 1.72 years) from Université de Moncton who volunteered to participate in exchange for course credits or an entry in a monthly $100 draw. Participants were native French speakers aged between 18 and 30 years with normal or corrected vision. This study was approved by the Ethics Board for Research Involving Humans of Université de Moncton.

Materials

Stimuli were 2,720 French nouns from the Lexique 3.83 database (New et al., 2004). All words had 8 letters and 2 syllables, with a frequency ranging from 0 to 739.12 occurrences per million (*M* = 7.49, *SD* = 31.36). For each participant, we sampled without replacement from this word pool to create 26 pure lists of 40 target words and 26 lists of 40 distractors, with 2 lists from each group serving as practice trials and 24 as experimental trials. Therefore, each participant was presented with 26 different lists of target words and distractors.

Design and Procedure

We used a 2 X 2 X 2 repeated measures design with production modality (produced vs. silent), presentation order at test (direct vs. random), and serial position (early positions vs. late positions) as within-subject factors. The experiment was controlled with PsyToolkit (Stoet, 2010, 2017) and stimuli were displayed in blue or white against a black background. Participants were tested individually in a quiet laboratory room and in a single session lasting approximately 90 minutes. Participants undertook two practice trials and 24 experimental trials (12 per production modality). During each trial, 40 to-be-remembered words were presented in the center of the screen at a rate of 2s per word (2,000 ms on, 0 ms off). Production modality was manipulated by creating two blocks of 12 pure lists and by counterbalancing their order across participants. Participants were instructed to read the blue words aloud and to read the white words silently, without mouthing or whispering the words.

Two seconds after the last word of each list was presented, a digit appeared in the center of the screen to signal the beginning of a 30s parity judgment task. During this task, a series of integers ranging from 0 to 9 appeared individually in the center of the screen. Participants were instructed to press the "Z" key if the stimulus was an odd number, and to press the "M" key if the stimulus was an even number. This task was self-paced, but lasted 30s for all participants. Immediately after the parity judgment task, participants completed a two-alternative forced choice (2AFC) recognition test. During this test, two words were presented next to each other in the center of the screen. Of the two words, one was part of the to-be-remembered list (target word), and the other was a distractor. Participants were instructed to click on the target word, and this task was then repeated for each of the 40 target words. Target words appeared equally often on either side of the screen. Finally, the presentation order of the target words at test was also counterbalanced, so that they were presented in the order in which they were encoded for half of the trials (direct order) and in a random order for the remaining trials.

Data Availability and Analysis

The stimuli for this experiment, the data for all experiments and the scripts used for data analyses are available on the Open Science Framework project page (OSF). Importantly, the data analysis procedure presented below was followed for all experiments. Statistical analyses were conducted using the *R* software (*R* Core Team, 2023). We collected descriptive information with frequentist analyses and guided our inferences with Bayesian analyses. Frequentist analyses were computed using the "lsr" package (version 0.5; Navarro, 2015), and Bayesian inferences were driven by Bayes Factor (BF) ANOVA analyses conducted using the "BayesFactor" package with the default parameters (Version 0.9.12-4.2; Morey & Rouder, 2018; Rouder et al., 2009, 2012). All Bayes Factors (BFs) were estimated via Monte Carlo simulations using 100,000 iterations, and when necessary, more iterations were added to reduce the proportional error below 5%.

Participants were entered as a random factor in each model, and we tested each main effect and interaction by successively omitting them from the full model. To report our results, we used BF_{10} to report evidence in favor of an effect (H1) and BF_{01} ($1/BF_{10}$) to report evidence against an effect (H0). Finally, based on guidelines by Goss-Sampson (2020), BF_{10} and BF_{01} values ranging from 1 to 3 were interpreted as anecdotal evidence, while values between 3 and 10, 10-30, 30-100 and over 100 were interpreted as moderate, strong, very strong and decisive evidence.

Results

Parity Judgment Task

To determine if participants were adequately engaged in the parity judgment task, we looked at the number of parity judgment attempts and the proportion of correct judgments as a function of production modality and presentation order at test. As can be seen in Table 1, participants were actively engaged in the task, and their performance did not differ across conditions. Results from Bayesian and Welch's paired *t*-tests also support this claim. For the number of attempts, there was moderate evidence against a difference between production modalities, $t(93.949) = 0.33$, $p = .74$, Cohen's $d = 0.07$, $BF_{01} = 4.50$, and anecdotal evidence against a difference between presentation orders, $t(93.919) = -0.22$, $p = .83$, Cohen's $d = 0.05$, $BF_{01} = 2.33$. For the proportion of correct judgments, there was anecdotal evidence against a difference between both production modalities, $t(87.244) = 0.86$, $p = .39$, Cohen's $d = 0.18$, $BF_{01} = 1.22$, and presentation orders, $t(93.463) = 0.64$, $p = .52$, Cohen's $d = 0.13$, $BF_{01} = 1.10$.

Two-Alternative Recognition Test

Mean recognition as a function of production modality, presentation order at test and serial position is shown in Figure 2. To facilitate a direct comparison with Fawcett et al. (2023), we reduced the 40-level serial position variable to a 2-level factor by calculating the average performance at the first and last three positions. We computed a 2 X 2 X 2 repeated measures

ANOVA with production modality (produced vs. silent), presentation order (direct vs. random) and serial position block (early vs. late) as factors (see Figure 3 for means). We found decisive evidence in favor of an effect of production modality, $F(1, 47) = 59.09$, $p < .001$, $\eta_p^2 = .56$, $BF_{10} > 10{,}000$, moderate evidence against an effect of presentation order, $F(1, 47) = 3.08$, $p = .08$, $\eta_p^2 = .06$, $BF_{01} = 3.37$, and moderate evidence against an effect of serial position block, $F(1, 47) = 2.08$, $p = .16$, $\eta_p^2 = .04$, $BF_{01} = 8.99$. We also found moderate evidence against the three-way interaction between production modality, presentation order and serial position block, $F(1, 47) = 1.26$, $p = .27$, $\eta_p^2 = .03$, $BF_{01} = 5.90$. For two-way interactions, we found strong evidence against an interaction between production modality and presentation order, $F(1, 47) = 0.07$, $p = .79$, $\eta_p^2 < .01$, $BF_{01} = 14.72$, and decisive evidence in favor of an interaction between presentation order and serial position block, $F(1, 47) = 21.45$, $p < .001$, $\eta_p^2 = .31$, $BF_{10} > 10{,}000$; as expected, the primacy effect was larger with the direct order.

Critically, although the interaction between production modality and serial position block was significant with the frequentist approach, the Bayesian ANOVA showed only anecdotal evidence, $F(1, 47) = 6.03$, $p < .05$, $\eta_p^2 = .11$, $BF_{10} = 1.69$. However, to allow for comparisons with results from Fawcett et al. (2023), we investigated the interaction by conducting a single factor ANOVA for each block (early vs. late positions) with production modality as the within-subject factor. Although the production effect was larger at the last serial positions, we found decisive evidence in favor of an effect of production modality (produced > silent) in both the early position block, $F(1, 47) = 19.63$, $p < .001$, $\eta_p^2 = .30$, $BF_{10} > 10{,}000$, and the late position block, $F(1, 47) = 60.64$, $p < .001$, $\eta_p^2 = .56$, $BF_{10} > 10{,}000$.

Exploratory Serial Position Analysis

The above-mentioned analyses were computed to replicate the procedure used by Fawcett et al. (2023). However, because most of the serial position curve was ignored in the analysis, it may not provide the most accurate picture. Therefore, we computed a logistic mixed-effects model. This analysis allowed the investigation of the interaction between the production effect and the entire serial position curve by considering serial position as an integer variable. The analyses were computed with the *glmr* function of the *lme4 R* package (Bates et al., 2015; *R* Core Team, 2023). We built the model by releasing the intercept along the participant and trial variables as random effects (see Guitard et al., 2023, for the importance of trials in memory), and including production modality, serial position, and presentation order as fixed effects. We found a significant main effect of production modality (produced > silent), $\chi^2(1) = 836.20$, $p < .001$, but no difference across presentation orders ($p = .38$) or serial positions ($p = .72$). Finally, as we observed with the Bayesian ANOVA, the only significant interaction was between presentation order and serial position (larger primacy effect in the direct order), $\chi^2(1) = 55.77$, $p < .001$. None of the other interactions reached significance ($p > .05$).

Discussion

In this experiment, we tested whether the cross-over interaction between the production effect and serial position observed in recall and order reconstruction tasks could be replicated with an item recognition task. We also tested whether output order at test would impact results. Overall, four findings emerged. First, we observed a large and reliable production effect across all serial positions. Second, we found that representing the items in their encoding order at test heightened the primacy effect but did not alter the size of the production effect. Third, while the production effect was numerically smaller at the first three positions compared to the last three positions, we found no evidence of the cross-over interaction between the production effect and serial position[1]. More specifically, the interaction between production and serial position block (early vs. late positions) reached significance with the frequentist approach, but it did not fit the cross-over pattern typically observed in recall tasks or hold up under Bayesian statistics. Finally, the similarity between our findings and those from Fawcett et al.'s (2023) meta-analysis further shows the robustness of the production effect. In fact, the studies reviewed by Fawcett et al. used a between-subjects design in which participants took part in the production condition or the silent reading condition. It could be argued that within-subjects designs might yield different results because participants could implement compensatory strategies driven by metamemory factors (Watkins et al., 2000). However, our results complement those of Gionet et al. (2022) in free recall by showing that this is not the case: The production effect is robust with both within- and between-subjects designs producing the same results with pure lists.

Experiment 2

The results of Experiment 1 revealed that the large memory cost for produced items typically observed in free recall tasks (Gionet et al., 2022) did not replicate in long-term item recognition, thus leading to the absence of a cross-over interaction between the production effect and serial position. How can we account for this important dissociation between recall and item recognition? The simplest hypothesis is to assume that both tasks highlight different memory processes. As suggested by MINERVA 2 (Jamieson et al., 2016) and the REM (Kelly et al., 2022; Wakeham-Lewis et al., 2022), producing the items in item recognition would add features at no cost, and as suggested by the attentional subsetting theory (Caplan & Guitard, 2024), it

[1] With 24 trials and a 30-second interference task per trial, it could be argued that results were due to a fatigue effect. To rule out this hypothesis, we recomputed the analyses by considering only the first block of trials (12 trials instead of 24) with the caveat that our design would be less powerful. In fact, production becomes a between-subjects factor instead of a within-subject factor. The main finding remained with no detrimental effect of production on the first serial positions. The complete analysis can be found on the Open Science Framework project page.

would add them at a constant cost across serial positions. On the other hand, as suggested by the RFM (Saint-Aubin et al., 2021), producing the items in recall tasks would interfere with their rehearsal. Therefore, the dissociation might be explained by a less important role of rehearsal in recognition than in recall, but this view can rapidly be dismissed because it is well-established that rehearsal also plays a critical role in recognition (Glenberg et al., 1977; Greene, 1987; Rundus et al., 1970; Woodward et al., 1973). In addition, Johnson and Miles (2009) showed that articulatory suppression, which blocks rehearsal, hindered performance on a 2AFC test.

Importantly, the absence of rehearsal related effects in our first experiment may not have been due to the memory task used, but rather to list length. In free recall, the proportion of words rehearsed largely decreases as list length increases (see, Bhatarah et al., 2009). Thus, according to the RFM, in the free recall of long lists, the cost of production on the first serial positions should be reduced or eliminated. As predicted, with 24-item lists and a free recall task, Cyr et al. (2022) observed a much-reduced disadvantage for produced items at the first serial positions with a large and reliable advantage at the last serial positions. This suggests that the differences between item recognition and recall observed in previous studies on the production effect may simply reflect the lower involvement of rehearsal with long lists rather than fundamental differences in the memory processes involved in both tasks.

Therefore, our second experiment was aimed at determining whether using short lists would allow us to uncover the cross-over interaction between the production effect and serial position. This test was implemented by using a set of pronounceable non-words created by Ward et al. (2005) to investigate the functional equivalence between order reconstruction and short-term recognition with a 2AFC test. We used non-words instead of words to ensure that performance would be at a reasonable level despite the very short lists used for the 2AFC memory test. In addition, all accounts of the production effect suggest that the same processes should be involved with words and non-words, resulting in the same pattern of results (e.g., Caplan & Guitard, 2024; Jamieson et al., 2016; Kelly et al., 2022; Saint-Aubin et al., 2021; Wakeham-Lewis et al., 2022). Finally, we removed the parity judgement task to allow direct comparison with performance in the immediate serial recall task used in Experiment 4.

Method

Participants

Forty-eight participants (30 women, 16 men, and 2 non-binary individuals) aged between 18 and 30 years (*M* = 25.35 years, *SD* = 2.72 years) were recruited through the Prolific online platform (https://www.prolific.com/). To be included in this study, participants also had to be native English speakers, to be from Canada, the United States, or the United Kingdom, to have normal or corrected vision, to have a Prolific approval rate of 90% or above, and to not have any language related disorders, cognitive impairments, or dementia. Participants were paid £4.50 for completing the study that lasted approximately 20 minutes. In addition, to ensure compliance with the instructions (specifically for the production condition), participants were asked to record and submit a video of themselves completing the study in order to receive compensation. In total, 10 participants were rejected for not providing a video or not following the instructions.

Materials

The stimuli used in Experiments 2 and 3 were a large set of 939 non-words developed by Ward et al. (2005). Each non-word was pronounceable and was constructed by combining two consonant clusters with a central vowel sound (e.g., “blove”, “jeeb”, “drard”). In addition, Ward et al. created two similar-sounding and unique distractors for each target non-word by changing either the initial or the final consonant cluster. For instance, the two distractors for “garch” were “karch” and “garge”. For each participant, we sampled without replacement from the original list of 939 target non-words to create 28 lists of 6 non-words, including 4 lists that served as practice trials and 24 lists that served as experimental trials. For the 2AFC memory test, the distractor for each target non-word was randomly selected from the two available options. Each participant was presented with 28 different lists of target non-words and distractors.

Design and Procedure

We used a 2 X 2 X 6 repeated measures design with production modality (produced vs. silent), presentation order at test (direct vs. random) and serial position (1 – 6) as within-subject factors. This experiment was conducted online, and participants were asked to record themselves while they completed the study. Videos were then reviewed to ensure task compliance. During the study, participants completed 4 practice trials and 24 experimental trials (12 per production modality). Unlike in Experiment 1 where production modality was blocked, pure lists of 6 non-words to be read aloud or silently were presented in a random order for each participant. Target non-words were presented individually at a rate of 2s per item (2,000 ms on, 0 ms off), and contrary to our first experiment where a 30-second parity judgment task was used, participants completed a 2AFC recognition test immediately after the last target was presented. Once again, target items appeared equally often on either side of the screen during the test, and their presentation order at test was counterbalanced (direct or random order).

Results

Analyses with all Serial Positions

First, we computed a 2 X 2 X 6 repeated measures ANOVA with production modality (produced vs. silent), presentation order (direct vs. random), and serial position (1 – 6) as factors (see Figure 4). We found decisive evidence in favor of an effect of production modality, $F(1, 47) = 121.53$, $p < .001$, $\eta_p^2 = .72$, $BF_{10} > 10{,}000$, strong evidence against an effect of presentation order, $F(1, 47) = 1.99$, $p = .17$, $\eta_p^2 = .04$, $BF_{01} = 12.46$, and decisive evidence against an effect of serial position, $F(5, 235) = 2.24$, $p = .05$, $\eta_p^2 = .05$, $BF_{01} = 792.45$. For two-way interactions, we found strong evidence against an interaction between production modality and presentation order, $F(1, 47) = 0.09$, $p = .76$, $\eta_p^2 = .002$, $BF_{01} = 24.90$, decisive evidence against an interaction between production modality and serial position, $F(5, 235) = 2.17$, $p = .06$, $\eta_p^2 = .04$, $BF_{01} = 290.62$, and decisive evidence against an interaction between serial position and presentation order, $F(5, 235) = 0.36$, $p = .88$, $\eta_p^2 = .007$, $BF_{01} > 10{,}000$. Finally, we also found decisive evidence against the three-way interaction between production modality, serial position and presentation order, $F(5, 235) = 1.37$, $p = .24$, $\eta_p^2 = .03$, $BF_{01} = 612.22$.

Analyses with Early and Late Serial Positions

Echoing our procedure from Experiment 1, we enabled comparisons between our results and those of Fawcett et al. (2023) by averaging performance at the first and last three positions. We conducted a 2 X 2 X 2 repeated measures ANOVA with production modality (produced vs. silent), presentation order (direct vs. random) and serial position block (early vs. late positions) as factors (see Figure 5). We found decisive evidence in favor of an effect of production modality, $F(1, 47) = 121.53$, $p < .01$, $\eta_p^2 = .72$, $BF_{10} > 10{,}000$, strong evidence against an effect of presentation order, $F(1, 47) = 1.99$, $p = .17$, $\eta_p^2 = .04$, $BF_{01} = 11.98$, and strong evidence against an effect of serial position block, $F(1, 47) = 1.94$, $p = .17$, $\eta_p^2 = .04$, $BF_{01} = 10.53$. For two-way interactions, there was strong evidence against the interaction between production modality and presentation order, $F(1, 47) = 0.09$, $p = .76$, $\eta_p^2 = .002$, $BF_{01} = 25.15$, very strong evidence against the interaction between serial position block and presentation order, $F(1, 47) = 0.20$, $p = .66$, $\eta_p^2 = .004$, $BF_{01} = 23.13$, and moderate evidence against the interaction between production modality and serial position block, $F(1, 47) = 2.14$, $p = .15$, $\eta_p^2 = .04$, $BF_{01} = 6.38$. Finally, we also found moderate evidence against the three-way interaction between production modality, presentation order and serial position block, $F(1, 47) = 1.73$, $p = .19$, $\eta_p^2 = .04$, $BF_{01} = 8.44$.

Discussion

Experiment 2 was aimed at testing whether using key task elements like those typically used in recall tasks would reveal a cross-over interaction between the production effect and serial position in item recognition. Contrary to our predictions, despite using short lists of 6 non-words and removing the filled retention interval, the cross-over interaction observed in recall tasks and illustrated in Figure 1 did not appear in item recognition. We observed a production effect across all serial positions, which fits well with MacLeod et al.'s (2022) findings who also observed a production effect with non-words. However, while the production effect appeared smaller at the first positions compared to the last positions, thus replicating our findings from Experiment 1 and those of Fawcett et al. (2023), we did not find evidence in favor of an interaction. More specifically, analyzing our data by considering all serial positions and by combining the first and last three positions respectively showed decisive and moderate evidence against an interaction between the production effect and serial position. These results suggest that the dissociation between recall and item recognition tasks, marked by the absence of a cross-over interaction in item recognition, is not due to key methodological differences beyond the retrieval requirements. Instead, it would seem that the memory processes underlying the cross-over interaction between the production effect and serial position in recall and order reconstruction tasks are differently impacted by the specific demands associated with an item recognition task.

Experiment 3

Results from our first two experiments show that although there is a trend towards a larger production effect at the last serial positions compared to the first serial positions, the cross-over interaction typically observed in free recall (Cyr et al., 2022; Gionet et al., 2022), immediate serial recall (Dauphinee et al., 2024; Macken et al., 2016) and order reconstruction (Jonker et al., 2014; Saint-Aubin et al., 2021) is not present in item recognition. However, before drawing any firm conclusions, one more key factor should be examined. In the first two experiments, memory was assessed with a 2AFC recognition test, and while this kind of test has been used to assess serial position effects in item recognition tasks (e.g., Johnson & Miles, 2009; Ward et al., 2005), it is less commonly used to investigate the production effect (see, Fawcett et al., 2022; Hopkins & Edwards, 1972; MacLeod et al., 2010). In fact, most item recognition studies on the production effect used old-new recognition tests (see, e.g., Bodner & Taikh, 2012; Bodner et al., 2014; Kelly et al., 2024; Ozubko & MacLeod, 2010).

Another argument for the use of an old-new recognition test was made by Yonelinas et al. (1992), who found that the list-strength effect (i.e., the lower memory for items within a list following the strengthening of other items from the same list) was observed in an old-new recognition test, but not in a 2AFC test. In addition, they also found that memory performance was generally greater on the old-new recognition test. To explain this, they suggested that participants perceive the old-new recognition test as more difficult, leading them to devote more effort to item encoding and resulting in more efficient encoding on an old-new recognition test. This is important, because according to the RFM (Cyr et al., 2022; Saint-Aubin et al., 2021), the cross-over interaction between the production effect and serial position in recall is driven by memory processes operating during item encoding. Therefore, using an old-new recognition test instead of a 2AFC test could maximize our chances of observing the same cross-over interaction between production and serial position in item recognition.

More recently, Brady et al. (2023) discussed differences between forced choice and old-new recognition tests (see also, Ulatowska et al., 2020). While they recognized the widespread use of old-new recognition tests in working memory research and strongly suggested using 2AFC tests instead, they also acknowledged that both tasks might yield different estimations of the memory strength distribution. In fact, by asking participants to choose which of two items in a

2AFC is old, we can determine which items have the stronger memory traces. Alternatively, by asking them to decide whether a single item is old or new, we can determine the items whose memory traces are strong enough to be recognized as old (Brady et al., 2023). This means that conclusions we make about memory performance, and particularly about differences across study conditions, may largely depend on the task used to assess memory in the first place. Therefore, to stay in line with previous studies on the production effect and to ensure that our results are not only due to the use of a 2AFC recognition test, Experiment 3 was aimed at replicating our results from Experiment 2 with an old-new recognition test.

Method

Participants

We recruited a second sample of 48 participants (28 women, 20 men, *M*age = 25.88 years, *SD* = 3.38 years) from the Prolific platform (https://www.prolific.com/). Inclusion and exclusion criteria were the same as for Experiment 2, and videos were collected from each participant and reviewed to ensure task compliance. Participants were paid £4.50. In total, 6 participants were rejected for not submitting videos or not following the instructions.

Materials, Design and Procedure

The stimuli, design and procedure were almost identical to those used in Experiment 2, except for the following changes. First, we simplified the design to a 2 X 6 repeated measures design with production modality (produced vs. silent) and serial position (1 – 6) as within-subject factors. Second, instead of a 2AFC recognition test, we used an old-new recognition test. After each list of target non-words was presented, participants completed a 12-item recognition test during which the 6 target non-words and 6 distractors (one similar-sounding distractor per target) were presented one at a time in a random order. Using their computer mouse, participants were asked to click on the "old" response if they believed the non-word had been studied and on the "new" response if they believed the non-word had not been studied.

Results

Analyses with all Serial Positions

We computed separate 2 X 6 repeated measure ANOVAs for hits, false alarms, and a signal detection measure of discrimination (*d*') with production modality (produced vs. silent) and serial position (1 – 6) as factors (see Figure 6 for the means). For hits, we found decisive evidence in favor of an effect of production modality, $F(1, 47) = 67.12$, $p < .001$, $\eta_p^2 = .59$, $BF_{10} > 10{,}000$, and in favor of an effect of serial position, $F(5, 235) = 12.37$, $p < .001$, $\eta_p^2 = .21$, $BF_{10} > 10{,}000$. However, there was decisive evidence against an interaction between production modality and serial position, $F(5, 235) = 1.75$, $p = .12$, $\eta_p^2 = .04$, $BF_{01} = 528.14$. For false alarms, there was moderate evidence in favor of both an effect of production modality, $F(1, 47) = 9.22$, $p < .01$, $\eta_p^2 = .16$, $BF_{10} = 8.39$, and an effect of serial position, $F(5, 235) = 4.87$, $p < .001$, $\eta_p^2 = .09$, $BF_{10} = 3.98$, but decisive evidence against the interaction between production modality and serial position, $F(5, 235) = 1.49$, $p = .19$, $\eta_p^2 = .03$, $BF_{01} = 786.53$. Finally, for *d*', there was decisive evidence in favor of an effect of production modality, $F(1, 47) = 24.18$, $p < .001$, $\eta_p^2 = .34$, $BF_{10} > 10{,}000$, moderate evidence against an effect of serial position, $F(5, 235) = 3.96$, $p < .01$, $\eta_p^2 = .08$, $BF_{01} = 3.53$, and very strong evidence against the interaction between production modality and serial position, $F(5, 235) = 1.13$, $p = .34$, $\eta_p^2 = .02$, $BF_{01} = 36.68$.

Analyses with Early and Late Serial Positions

As in our previous experiments, we transformed the 6-level serial position variable into a 2-level factor for hits and *d*' by averaging performance at the first and last three serial positions. Then, we computed separate 2 X 2 repeated measure ANOVAs for each dependent variable with production modality (produced vs. silent) and serial position block (early vs. late) as factors (see

Figure 7 for the means). For hits, there was decisive evidence in favor of an effect of production modality, $F(1, 47) = 67.12$, $p < .001$, $\eta_p^2 = .59$, $BF_{10} > 10{,}000$, and decisive evidence in favor of an effect of serial position block, $F(1, 47) = 30.59$, $p < .001$, $\eta_p^2 = .39$, $BF_{10} > 10{,}000$. However, there was only anecdotal evidence in favor or against an interaction between production modality and serial position, $F(1, 47) = 4.37$, $p = .04$, $\eta_p^2 = .09$, $BF_{01} = 2.40$. For *d*', there was decisive evidence in favor of an effect of production modality, $F(1, 47) = 24.18$, $p < .001$, $\eta_p^2 = .34$, $BF_{10} > 10{,}000$, but moderate evidence against an effect of serial position block, $F(1, 47) = 3.86$, $p = .06$, $\eta_p^2 = .08$, $BF_{01} = 3.01$, and against the interaction between production modality and serial position block, $F(1, 47) = 2.27$, $p = .14$, $\eta_p^2 = .05$, $BF_{01} = 3.30$.

Discussion

Results from this experiment nicely extended those observed in Experiment 2. Overall recognition performance was slightly reduced compared to our previous two experiments, which is in slight contrast with results from Yonelinas et al. (1992) who found that performance was generally higher in an old-new recognition test compared to a 2AFC. Despite this small overall difference in performance, our results once again show a numerically larger benefit for produced items at the last serial positions. Specifically, the analyses based on all serial positions revealed very strong and decisive evidence against the interaction. However, additional analyses based on the early and late serial position blocks were more nuanced, showing anecdotal and moderate evidence against said interaction. Overall, these results suggest that the pattern observed in Experiment 2 was not due to the use of a 2AFC, and that with short lists, regardless of the recognition test used, the production effect does not interact with serial position the same way that it does in recall.

Experiment 4

Results from our last two experiments nicely replicated the overall pattern observed in Experiment 1. Thus far, we can conclude that regardless of list length or the recognition test used, although the production effect is larger at the last serial positions compared to the first positions, the cross-over interaction between the production effect and serial position observed in recall is not present in item recognition. However, it could be argued that a change in stimuli from words to non-words prevented us from making any further comparisons between our results and what is typically observed in recall tasks. The production effect for non-words has been studied with old-new recognition tasks (e.g., MacLeod et al., 2010; 2022) and with various implicit memory tasks (see, Lu & MacLeod, this issue; Mama, this issue). However, to our knowledge, it has never been studied with recall tasks. Therefore, it is possible that in recall tasks, the production effect with non-words does not interact with serial position. If this happens, item recognition and recall tasks would be more similar than different.

According to the Revised Feature Model (see, Cyr et al., 2022; Saint-Aubin et al., 2021), the production effect in verbal recall tasks should remain consistent across stimuli variations to the extent that they are pronounceable. That is, the last produced non-words of the list should still benefit from enhanced distinctiveness due to the additional features generated by production. In addition, we should still observe a cost for the first produced non-words, whose rehearsal should be blocked by production. Unlike what we observed with item recognition tasks in the previous experiments, this should result in the absence of an overall production effect and in a cross-over interaction between the production effect and serial position. Therefore, Experiment 4 was aimed at testing the production effect with non-words and an immediate serial recall task to determine the generalizability of the cross-over interaction between the production effect and serial position.

Method

Participants

We recruited a sample of 48 participants (32 women, 12 men, 4 non-binary individuals, *M*age = 25.08, *SD* = 3.04) from the Prolific platform (https://www.prolific.com/) by using the same criteria as our previous 2 experiments. We also assessed task compliance by collecting videos from each participant. Participants were paid £4.50, and 14 participants were excluded for not submitting videos or following task instructions (e.g., not reading the blue non-words aloud).

Materials, Design and Procedure

The stimuli, design and procedure were identical to those used in Experiment 3, except for the memory task. After the presentation of each list of 6 non-words, participants had to complete an immediate serial recall task in which they had to recall as many non-words as possible from the last encoded list. They were also instructed to recall the non-words in their presentation order, from the first non-word to the last. Using their keyboard, participants typed the non-words and pressed the enter key after each non-word to register their answer. Then, the recalled non-word, if any, left the screen and participants were prompted to recall the following non-word. Participants were also able to skip an item by pressing the enter key, but could not return to edit any previous answers. Finally, when participants were done recalling the items, they pressed the space bar to start the next trial. This procedure was then repeated for all 28 trials.

Data Analysis and Results

Prior to conducting any data analyses, we checked participants' answers for misspelling and corrected all responses that could be identified without ambiguity (e.g., letter repetitions: sisske instead of siske; letter omissions: sike instead of siske; or letter substitutions: simke instead of siske). All analyses reported below were based on corrected data, and while correcting spelling errors marginally increased the overall performance, it had no effect on the pattern of results.

We conducted a 2 X 6 repeated measures ANOVA with production modality (produced vs. silent) and serial position (1-6) as factors (see Figure 8 for the means). There was very strong evidence against an effect of production modality, $F(1, 47) = 0.01$, $p = .91$, $\eta_p^2 < .001$, $BF_{01} = 36.18$, but decisive evidence in favor of an effect of serial position, $F(5, 235) = 16.10$, $p < .001$, $\eta_p^2 = .26$, $BF_{10} > 10{,}000$, and decisive evidence in favor of the interaction between production modality and serial position, $F(5, 235) = 22.79$, $p < .001$, $\eta_p^2 = .33$, $BF_{10} > 10{,}000$. In addition, we conducted a 2 X 2 repeated measures ANOVA with production modality and serial position block (early vs. late) as factors (see Figure 9). Once again, there was very strong evidence against an effect of production modality, $F(1, 47) = 0.01$, $p = .91$, $\eta_p^2 < .001$, $BF_{01} = 36.62$, along with decisive evidence in favor of an effect of serial position block, $F(1, 47) = 3.08$, $p = .09$, $\eta_p^2 = .06$, $BF_{10} > 6{,}050.34$, and decisive evidence in favor of the interaction between production modality and serial position block, $F(1, 47) = 44.39$, $p < .001$, $\eta_p^2 = .49$, $BF_{10} > 10{,}000$.

Discussion

In this experiment, we tested whether the cross-over interaction between the production effect and serial position typically observed in recall tasks (see, Dauphinee et al., 2024; Gionet et al., 2022) could be extended to non-word stimuli. As predicted, our results nicely replicated the cross-over interaction, with an advantage for produced items at the last position and a cost at the first positions resulting in the absence of an overall production effect. These results fit well with the Revised Feature Model (e.g., Cyr et al., 2022; Saint-Aubin et al., 2021), suggesting that the processes underlying this interaction between the production effect and serial position in recall are consistent across stimulus types. Most importantly, these results also provide critical evidence showing that our results from Experiments 2 and 3 were not due to the use of non-word stimuli and highlight the differential impact of task-specific processes on the interaction between the production effect and serial position.

Combined Analyses

Results from our first three experiments revealed a highly consistent and robust pattern: Across list lengths, recognition tasks and stimulus types, the production effect in item recognition is larger at the last serial positions compared to the first positions. However, our analyses of the interaction between production modality and serial position in item recognition had inconsistent outcomes. More specifically, in Experiment 1, the interaction was significant with a frequentist approach, but the BF ANOVA showed only anecdotal evidence. In Experiment 2, the interaction with serial position (1-6) nearly reached significance with the frequentist approach (p = .06), but the BF ANOVA showed evidence favoring the null (H0). Finally, in Experiment 3, the interaction with serial position block (early vs. late) was significant with hits as the dependent variable (BF ANOVA still showed anecdotal evidence), but not with *d*'. Based on these results, we decided to conduct additional analyses by combining data of Experiments 1-3 in order to maximize power to detect the interaction between production and serial position (see Figure 10).

Echoing the meta-analytic approach used by Fawcett et al. (2023) to combine data from studies with different list lengths, we averaged recognition performance at the first and last three serial positions. In Experiments 1 and 2, we analyzed the direct and random presentation orders together, and in Experiment 3, we chose hits as our dependent variable so that all data would be on the same scale. Finally, we computed a 2 X 2 X 3 mixed ANOVA with production modality (produced vs. silent) and serial position block (early vs. late) as the within-subject factors, and experiment (1 – 3) as the between-subjects factor. There was decisive evidence in favor of an effect of production modality, $F(1, 141) = 220.58$, $p < .001$, $\eta_p^2 = .61$, $BF_{10} > 10,000$, decisive evidence in favor of an effect of serial position block, $F(1, 141) = 16.16$, $p < .001$, $\eta_p^2 = .10$, $BF_{10} = 1,703.03$, and very strong evidence in favor of an effect of experiment, $F(2, 141) = 11.02$, $p < .001$, $\eta_p^2 = .14$, $BF_{10} = 44.95$. For two-way interactions, there was decisive evidence against an interaction between production modality and experiment, $F(2, 141) = 1.09$, $p = .34$, $\eta_p^2 = .02$, $BF_{01} = 133.54$, decisive evidence in favor of an interaction between serial position block and experiment, $F(2, 141) = 16.20$, $p < .001$, $\eta_p^2 = .19$, $BF_{10} > 10{,}000$, and strong evidence in favor of an interaction between production modality and serial position block, $F(1, 141) = 11.85$, $p = .001$, $\eta_p^2 = .08$, $BF_{10} = 24.01$. Finally, there was decisive evidence against the three-way interaction between serial position block, production modality and experiment, $F(2, 141) = 0.18$, $p = .84$, $\eta_p^2 = .002$, $BF_{01} = 342.72$.

General Discussion

In their review paper, MacLeod and Bodner (2017) proposed a list of unresolved issues and opportunities. They began their list by mentioning that in the context of pure lists, "one important unresolved issue is why production improves recognition and cued recall but not free recall." (p. 393) We subsequently showed that the absence of an effect in free recall was more apparent than real: Production enhances recall of the last serial positions while hindering recall of the first serial positions (e.g., Cyr et al., 2022; Gionet et al., 2022). The interaction with serial position further shows the usefulness of serial positions as a benchmark phenomenon to inform working memory theories and models (see, Oberauer et al., 2018).

In this paper, we systematically tested the presence of a cross-over interaction between the production effect and serial position in item recognition. To achieve this, we tried to make the recognition task as similar as possible to recall tasks in which an interaction has been reported by including a condition in which the items were tested in their presentation order, by reducing list length to six items, and by removing the delay between presentation and test. We also used a 2AFC and an old-new recognition test to ensure that our results were not due to the peculiarities of a specific recognition task (e.g., Brady et al., 2023; Yonelinas et al., 1992). Then, we used an immediate serial recall task to ensure that the basic cross-over interaction between the production effect and serial position could be observed with non-word stimuli. Finally, we combined the data from all our recognition experiments and conducted further analyses to maximize the probability of observing the targeted interaction. Overall, results are clear: the production effect in recognition is larger at the last serial positions compared to the first positions, resulting in a small interaction that is functionally different from the cross-over interaction typically observed in recall and order reconstruction tasks. As such our results confirmed those of a recent meta-analysis (Fawcett et al., 2023) and highlighted the modulation of basic memory processes by specific task demands. In addition, our results are also supported by a supplementary analysis of other data within the current issue (see, Whitridge et al., 2024).

The similarities between the current results and those reported in Fawcett's et al. (2023) meta-analysis further show the robustness of the effect. In fact, while this study used a repeated-measures design, all previous item recognition studies included in Fawcett et al.'s serial position analysis used a between-subjects design in which one group read the items aloud and another read them silently. The similarity across designs nicely replicated the free recall results of Gionet et al. (2022), who extended the pattern previously observed with a repeated-measures design to a between-subjects design. Furthermore, the nine datasets used by Fawcett et al. (2023) came from five articles published by the same two laboratories and using very similar methodologies. Here, it is reassuring that the same results were not only observed with new data from an independent laboratory, but also remained consistent across list lengths, recognition tasks and stimulus types.

It is also worth mentioning that the current results with item recognition tests and verbal materials nicely echoed those observed with visual-spatial materials in a recall task. In this special issue, Saint-Aubin et al. (2024) sequentially presented a short series of dots at various locations on a screen. Participants produced the items by clicking on them at presentation. At recall, participants were asked to indicate the location of the dots in their presentation order.

Overall, there was a large production effect. Most importantly, as reported here, in some experiments, an interaction was observed, but it was not a crossover, and in others it was totally absent (e.g., Experiment 3A). Saint-Aubin et al. fit the data with the RFM while maintaining a differential rehearsal rate for produced and control items.

Overall, current results raised an important question: How can we account for the dissociation between recall and item recognition tasks in the verbal domain? As a reminder, in recall tasks, the detrimental effect of production on the first serial positions has been attributed to disrupted rehearsal (Cyr et al., 2022; Dauphinee et al., 2024; Gionet et al., 2022; Saint-Aubin et al., 2021). To test this hypothesis in immediate serial recall, Dauphinee et al. manipulated presentation rate under the assumption that fast presentation rates would not allow participants to rehearse (Bhatarah et al., 2009; Landry et al., 2022; Macken et al., 2016; Tan & Ward, 2000, 2008). As predicted, there was no cross-over interaction at 500 ms per item, but it was present at 1000 ms and 2000 ms per item. At 500 ms per item, there was an advantage for produced items compared to silently read items on the last serial positions, but no difference was observed on the first serial positions. Dauphinee et al.'s (2024) systematic review confirmed these conclusions.

In their conclusion, after having uncovered the cross-over interaction between the production effect and serial position, Saint-Aubin et al. (2021) mentioned that "further research is necessary to establish whether these patterns can also be found within paradigms that are more typical of episodic memory research." (p.24). They further suspected that any negative impact of production on rehearsal is likely to be complex because of how rehearsal interacts with many factors including recall order. This view was echoed by Ward (2022) who acknowledged that it is difficult to fully understand the functions of rehearsal without considering the specificity of the tasks in which it is assumed to be involved. For instance, it has been suggested that rehearsal can increase the strength of item representations in long-term memory tasks (e.g., Greene, 1987) or contribute to maintaining items in serial order in a highly accessible state in short-term recall tasks (e.g., Baddeley, 1986; Barrouillet, Bernardin, & Camos, 2004).

After reviewing data relevant to five long-term memory tasks, Greene (1987) concluded that the effects of rehearsal are more obvious in long-term recognition than in recall tasks. This conclusion was later modified by Hartshorne and Makovski (2019), suggesting that the beneficial effect of rehearsal was similar for recall and recognition tasks. Therefore, the differential impact of production as a function of serial position in recall and recognition tasks cannot be attributed to a lack of rehearsal in recognition. As suggested by Ward (2022), although rehearsal has a large impact in many tasks, it may arise for different reasons. For instance, in immediate serial recall and order reconstruction, the role of order information is self-evident. Similarly, the role of order information in free recall performance is well established (Beaman & Jones, 1998; Ward & Tan, 2023). Therefore, one possible explanation for the cross-over interaction in verbal recall tasks not replicating in item recognition would be the dual cost of production in recall tasks. Under this view, in recall, produced items would suffer more from the lack of rehearsal because saying the items aloud would disrupt both order information and item strengthening, while in recognition it would only disrupt item strengthening. This larger cost would account for the detrimental effect of production on the first serial positions in recall tasks and for the smaller production effect at the first serial positions compared to the last positions in item recognition.

Another possible account for the difference between recall and recognition would be to assume that participants do not rehearse the same way in both tasks. Under this view, rehearsal would be calling upon the same processes as those involved in retrieval (e.g., Laming, 2006; Ward, 2022). As Ward (2022) nicely put it: “this approach constrains the variety of different retrieval strategies available at rehearsal to the different retrieval strategies available at test and may set limits on the upper bounds of when serial rehearsal remains effective”. (p.51) Further research is needed to test this hypothesis. Critically, participants could be presented with a list of items and only informed at test whether it would be a recall or a recognition task. This strategy has successfully been used to highlight processing differences as a function of foreknowledge of recall conditions (Guitard et al., 2020). As of yet, results are compatible with all current models of the production effect in item recognition (see, Caplan & Guitard, 2024; Jamieson et al., 2016; Kelly et al., 2022; Wakeham-Lewis et al., 2022). Furthermore, although the RFM does not apply to item recognition, it has been able to account for recall of visual-spatial information interacting with serial position in a similar way to what is observed here (Saint-Aubin et al., 2024). The critical test for models would be to account for both recall and recognition at the same time.

Conclusion

The current study proposed the first comprehensive test of the interaction between production and serial position in recognition tests while using methods as similar as possible to recall. Over all three item recognition experiments, results revealed a large and reliable beneficial effect of production with pure lists, thus reproducing previous work in the field (Fawcett, 2013; Fawcett et al., 2023). This demonstration by itself is important because it had previously been argued that the production effect could not be observed with pure lists (e.g., Hourihan & MacLeod, 2008; MacLeod et al., 2010; Ozubko & MacLeod, 2010). Here, we showed that not only can a beneficial effect of production be observed with pure lists and a between-subjects design, but it can also be observed with a within-subject design. Most importantly, we showed that the interaction between production and serial position is functionally different in recall and item recognition tasks. Specifically, across list lengths, recognition tasks and stimulus types, we found that the production effect in item recognition is larger at the last serial positions compared to the first positions. In addition, after combing the data from all three recognition experiments, we were able to show, for the first time, clear evidence in favor of an interaction between the production effect and serial position in item recognition. Finally, after conducting an additional experiment in immediate serial recall, we highlighted the impact of task-related demands on the memory processes underlying this interaction between production and serial position. Overall, following Ward's (2022) suggestion, it appears that rehearsal processes impact performance in recall and recognition tasks differently. In fact, each task is likely to maximize these rehearsal processes differently. Once more, the production effect demonstrates its usefulness in uncovering key memory processes.

References

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General, 133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *The Quarterly Journal of Experimental Psychology: Section A, 51(3)*, 615-636. https://doi.org/10.1080/713755774

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*(5), 689-713. https://doi.org/10.3758/MC.37.5.689

Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1711–1719. https://doi.org/10.1037/a0028466

Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review, 21*(1), 149-154. https://doi.org/10.3758/s13423-013-0485-1

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review, 30*(1), 421-449. https://doi.org/10.3758/s13423-022-02179-w

Caplan, J. B., & Guitard, D. (2024). A feature-space theory of the production effect in recognition. *Experimental Psychology, 71*(1), 64-82. https://doi.org/10.1027/1618-3169/a000611

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Cyr, V., Poirier, M., Yearsley, J. M., Guitard, D., Harrigan, I., & Saint-Aubin, J. (2022). 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)

Dauphinee, I., Roy, M., Guitard, D., Yearsley, J. M., Poirier, M., & Saint-Aubin, J. (2024). Give me enough time to rehearse: presentation rate modulates the production effect. *Psychonomic Bulletin & Review*, 1-12. https://doi.org/10.3758/s13423-023-02437-5

Engle, R. W., Cantor, J., & Turner, M. L. (1989). Modality effects : Do they fall on deaf ears? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 41*(2-A), 273-292. https://doi.org/10.1080/14640748908402366

Ensor, T. M., Guitard, D., Bireta, T. J., Hockley, W. E., & Surprenant, A. M. (2020). The list-length effect occurs in cued recall with the retroactive design but not the proactive design. *Canadian Journal of Experimental Psychology, 74*(1), 12–24. https://doi.org/10.1037/cep0000187

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fawcett, J. M. (2013). The production effect benefits performance in between-subjects designs: A meta-analysis. *Acta Psychologica*, *142*(1), 1-5. https://doi.org/10.1016/j.actpsy.2012.10.001

Fawcett, J. M., Baldwin, M. M., Whitridge, J. W., Swab, M., Malayang, K., Hiscock, B., Drakes, D. H., & Willoughby, H. V. (2023). Production improves recognition and reduces intrusions in between-subject designs: An updated meta-analysis. *Canadian Journal of Experimental Psychology, 77*(1), 35–44. https://doi.org/10.1037/cep0000302

Fawcett, J. M., Bodner, G. E., Paulewicz, B., Rose, J., & Wakeham-Lewis, R. (2022). Production can enhance semantic encoding: Evidence from forced-choice recognition with homophone versus synonym lures. *Psychonomic Bulletin & Review, 29*(1), 2256-2263. https://doi.org/10.3758/s13423-022-02140-x

Forrin, N. D., Ralph, B. C., Dhaliwal, N. K., Smilek, D., & MacLeod, C. M. (2019). Wait for it… performance anticipation reduces recognition memory. *Journal of Memory and Language, 109*, 104050. https://doi.org/10.1016/j.jml.2019.104050

Gionet, S., Guitard, D., & Saint-Aubin, J. (2022). The production effect interacts with serial positions: Further evidence from a between-subjects manipulation. *Experimental Psychology, 69*(1), 12-22. https://doi.org/10.1027/1618-3169/a000540

Glenberg, A., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning & Verbal Behavior, 16(3)*, 339–352. https://doi.org/10.1016/S0022-5371(77)80055-8

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin, 102(3)*, 403–413. https://doi.org/10.1037/0033-2909.102.3.403

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67*(1), 106-148. https://doi.org/10.1016/j.jml.2012.04.004

Grenfell-Essam, R., Ward, G., & Tan, L. (2017). Common modality effects in immediate free recall and immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1909–1933. https://doi.org/10.1037/xlm0000430

Gross-Sampson, M. (2020). New Bayesian Guide. https://doi.org/10.17605/OSF.IO/CKNXM

Guitard, D., Neath, I., & Saint-Aubin, J. (2023). Additional evidence that valence does not affect serial recall. *Quarterly Journal of Experimental Psychology*, *76*(8), 1790-1796. https://doi.org/10.1177/17470218221126635

Guitard, D., Saint-Aubin, J., Poirier, M., Miller, L. M., & Tolan, A. (2020). Forward and backward recall: Different visuospatial processes when you know what's coming. *Memory & Cognition, 48(1)*, 111–126. https://doi.org/10.3758/s13421-019-00966-w

Hartshorne, J. K., & Makovski, T. (2019). The effect of working memory maintenance on long-term memory. *Memory & Cognition, 47*(1), 749-763. https://doi.org/10.3758/s13421-019-00908-6

Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation Effects in Recognition Memory. *Journal of Verbal Learning and Verbal Behavior*, *11*(4), 534-537. https://doi.org/10.1016/S0022-5371(72)80036-7

Hourihan, K., & Fawcett, J. M. (2024). It's all about the case: production and reading fluency. *Experimental Psychology.*

Hourihan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of*

Experimental Psychology / Revue canadienne de psychologie expérimentale, 62(4), 242–246. https://doi.org/10.1037/1196-1961.62.4.242

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology, 70(2)*, 154–164. https://doi.org/10.1037/cep0000081

Johnson, A. J., & Miles, C. (2009). Serial position effects in 2-alternative forced choice recognition: Functional equivalence across visual and auditory modalities. *Memory, 17(1)*, 84–91. https://doi.org/10.1080/09658210802557711

Jonker, T. R., Levene, M., & MacLeod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(2)*, 441–448. https://doi.org/10.1037/a0034977.

Kelly, M. O., Ensor, T. M., Lu, X., MacLeod, C. M., & Risko, E. F. (2022). Reducing retrieval time modulates the production effect: Empirical evidence and computational accounts. *Journal of Memory and Language, 123*, 104299. https://doi.org/10.1016/j.jml.2021.104299

Kelly, M. O., Lu, X., Ensor, T. M., MacLeod, C. M., & Risko, E. F. (2024). Productions need not match study items to confer a production advantage, but it helps. *Experimental Psychology, 71*(1), 2-13. https://doi.org/10.1027/1618-3169/a000600

Lambert, A. M., Bodner, G. E., & Taikh, A. (2016). The production effect in long-list recall: In no particular order? *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 70*(2), 165–176. https://doi.org/10.1037/cep0000086

Laming, D. (2006). Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(5)*, 1146–1163. https://doi.org/10.1037/0278-7393.32.5.1146

Landry, É. R., Guitard, D., & Saint-Aubin, J. (2022). Arousal affects short-term serial recall. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 76*(2), 99–110. https://doi.org/10.1037/cep0000272

Lu, X., & MacLeod, C. M. (2024). Does production influence implicit memory? *Experimental Psychology*.

Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition*, *155*, 113-124. https://doi.org/10.1016/j.cognition.2016.06.013

MacLeod, C. M., & Bodner, G. E. (2017). The Production Effect in Memory. *Current Directions in Psychological Science*, *26*(4), 390-395. https://doi.org/10.1177/0963721417691356

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 671–685. https://doi.org/10.1037/a0018785

MacLeod, C. M., Ozubko, J. D., Hourihan, K. L., & Major, J. C. (2022). The production effect is consistent over material variations: support for the distinctiveness account. *Memory, 30*(8), 1000-1007. https://doi.org/10.1080/09658211.2022.2069270

Mama, Y. (2024). An explicit production effect in implicit memory tasks: Testing implicit memory using the production effect paradigm supports contamination by explicit intrusions. *Experimental Psychology*.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morey, R. D. & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology*, *21*(3), 263–276. https://doi.org/10.1037/h0082978

Murray, D.J., Leung, C., & McVie, D.F. (1974). Vocalization, primary memory and secondary memory. *British Journal of Psychology*, *65*(3), 403-413. https://doi.org/10.1111/j.2044-8295.1974.tb01414.x

Navarro, D. J. (2015). Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.5) University of Adelaide. Adelaide, Australia

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

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

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1543–1547. https://doi.org/10.1037/a0020604

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factor for ANOVA designs. *Journal of Mathematical Psychology*, *56*(1), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*(1), 63–77. https://doi.org/10.1037/h0031185

Rundus, D., Loftus, G. R., & Atkinson, R. C. (1970). Immediate free recall and three-week delayed recognition. *Journal of Verbal Learning & Verbal Behavior, 9(6),* 684–688. https://doi.org/10.1016/S0022-5371(70)80033-0

Saint-Aubin, J., Poirier, M., Yearsley, J.M., & Guitard, D. (2024). The production effect becomes spatial. *Experimental Psychology, 71(1), 14-32.* https://doi.org/10.1027/1618-3169/a000609

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, 118.* https://doi.org/10.1016/j.jml.2021.104219

Schwell., R., Icht, M., Reznick, J., & Mama, Y. (2024). Exploring the production effect in memory reveals a balanced bilingual advantage. *Experimental Psychology, 71*(1), 51-63. https://doi.org/10.1027/1618-3169/a000613

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4(2),* 145–166. https://doi.org/10.3758/BF03209391

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

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1589–1625. https://doi.org/10.1037/0278-7393.26.6.1589

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*(3), 535-542. https://doi.org/10.3758/PBR.15.3.535

Ulatowska, J., Olszewska, J., & Hanson, M. D. (2020). Does Test Type Influence False Recognition in the DRM Paradigm? Comparison of the Yes/No Recognition Test and Two-Alternative Forced-Choice Test. *The American Journal of Psychology*, *133*(1), 49-62. https://doi.org/10.5406/amerjpsyc.133.1.0049

Wakeham-Lewis, R. M., Ozubko, J., & Fawcett, J. M. (2022). Characterizing production: the production effect is eliminated for unusual voices unless they are frequent at study. *Memory*, *30(10)*, 1319-1333, https://doi.org/10.1080/09658211.2022.2115075

Ward, G. (2022). Rehearsal processes. *Oxford Handbook of Human Memory*.

Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory : Functional equivalence across modalities. *Memory*, *13*(3-4), 308-317. https://doi.org/10.1080/09658210344000279

Ward, G., & Tan, L. (2023). The role of rehearsal and reminding in the recall of categorized word lists. *Cognitive Psychology*, *145*. https://doi.org/10.1016/j.cogpsych.2023.101563

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1207–1241. https://doi.org/10.1037/a0020122

Watkins, M. J., LeCompte, D. C., & Kim, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 239–245. https://doi.org/10.1037/0278-7393.26.1.239

Whitridge, J. W., Huff, M. J., Ozubko, J. D., Bürkner, P. C., Lahey, C. D., & Fawcett, J. M. (2024). Singing does not necessarily improve memory more than reading aloud: An empirical and meta-analytic investigation. *Experimental Psychology, 71*(1), 33-50. https://doi.org/10.1027/1618-3169/a000614

Woodward, A. E., Bjork, R. A., & Jongeward, R. H. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning & Verbal Behavior, 12(6)*, 608–617. https://doi.org/10.1016/S0022-5371(73)80040-4

Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 345-355. https://doi.org/10.1037/0278-7393.18.2.345

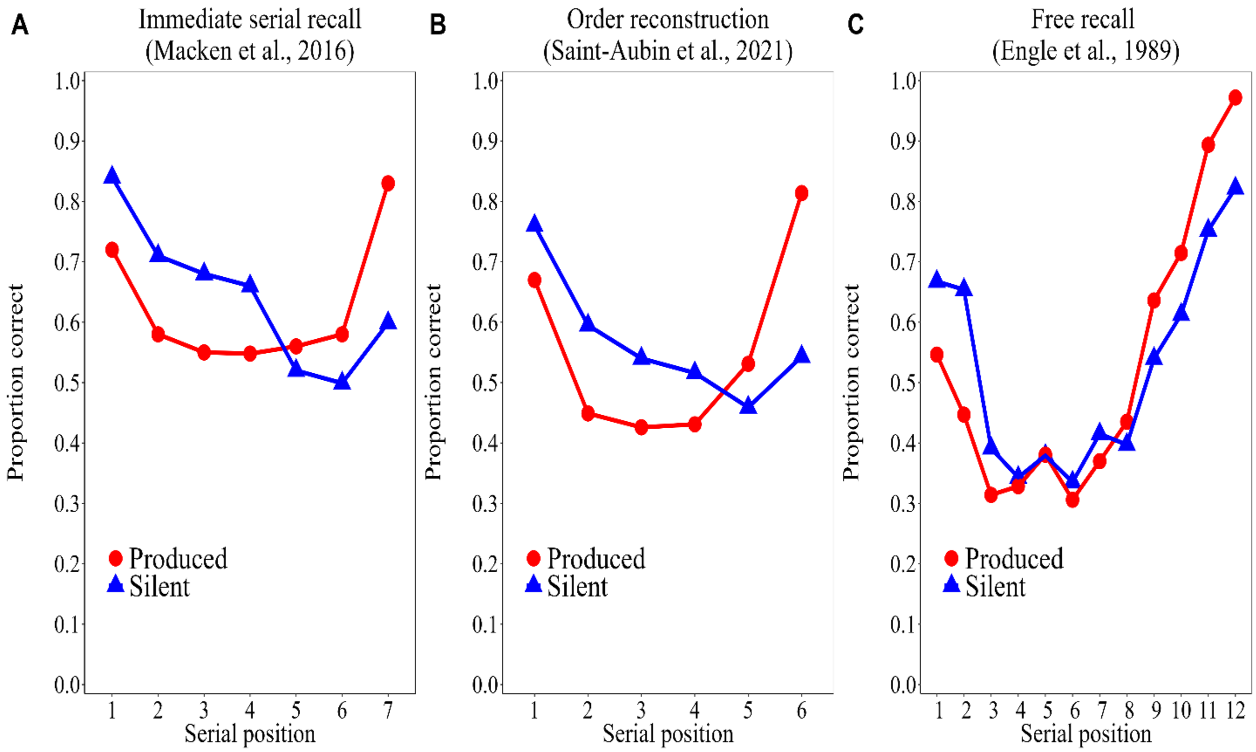
Table 1

Means and Standard Deviations (in Parentheses) for the Number of Parity Judgment Attempts and the Proportion of Correct Parity Judgments in Experiment 1.

Condition	Number of attempts	Proportion correct
Production modality		
Produced	43.55 (6.14)	0.97 (.02)
Silent	43.14 (6.29)	0.96 (.03)
Presentation order at test		
Direct order	43.21 (6.11)	0.97 (.02)
Random order	43.48 (5.93)	0.96 (.03)

Figure 1

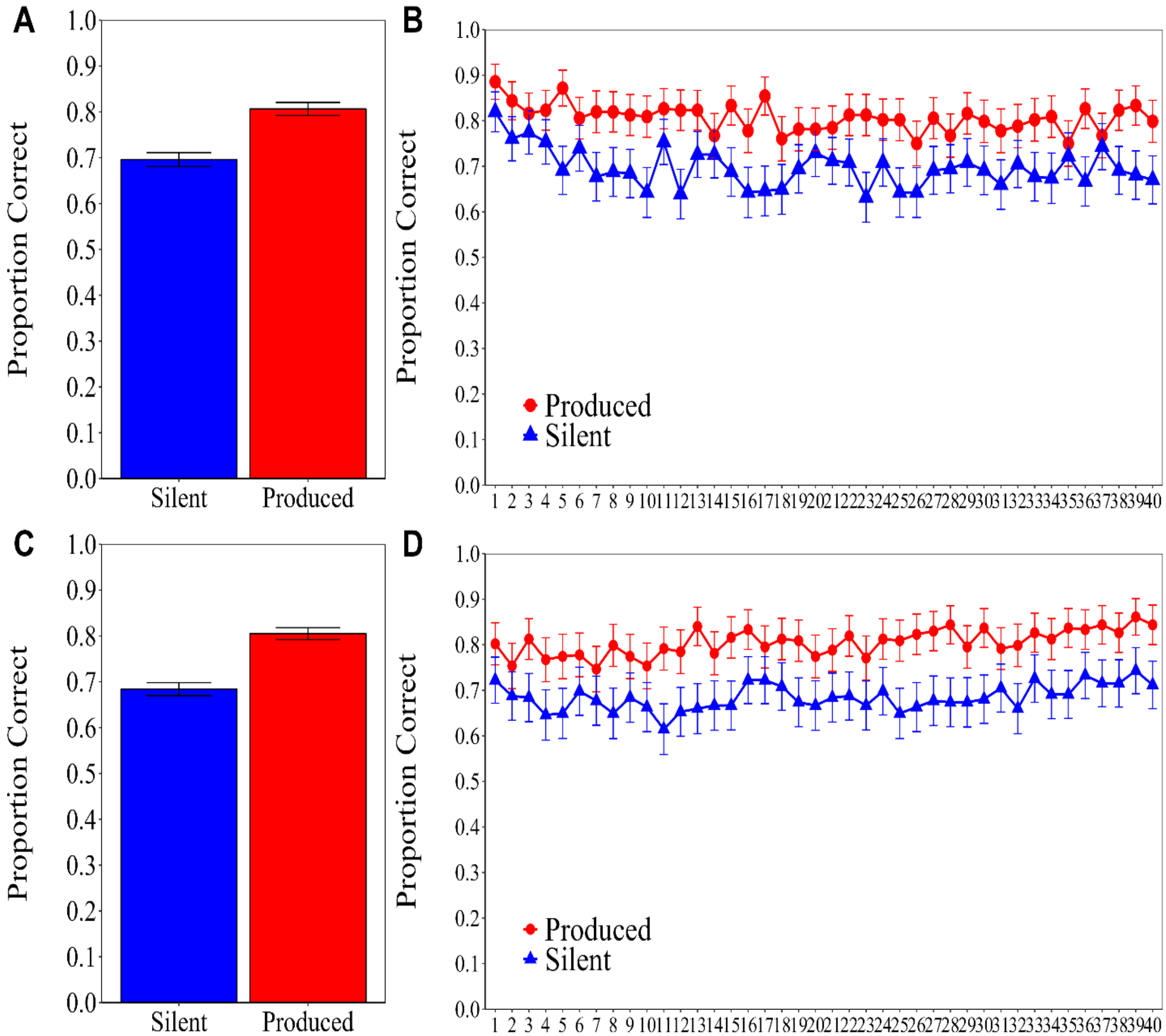
Performance for produced and silent items as a function of serial position and memory task.



Note. Panel A = Immediate serial recall (Macken et al., 2016), Panel B = Order reconstruction (Saint-Aubin et al., 2021), Panel C = Free recall (Engle et al., 1989).

Figure 2

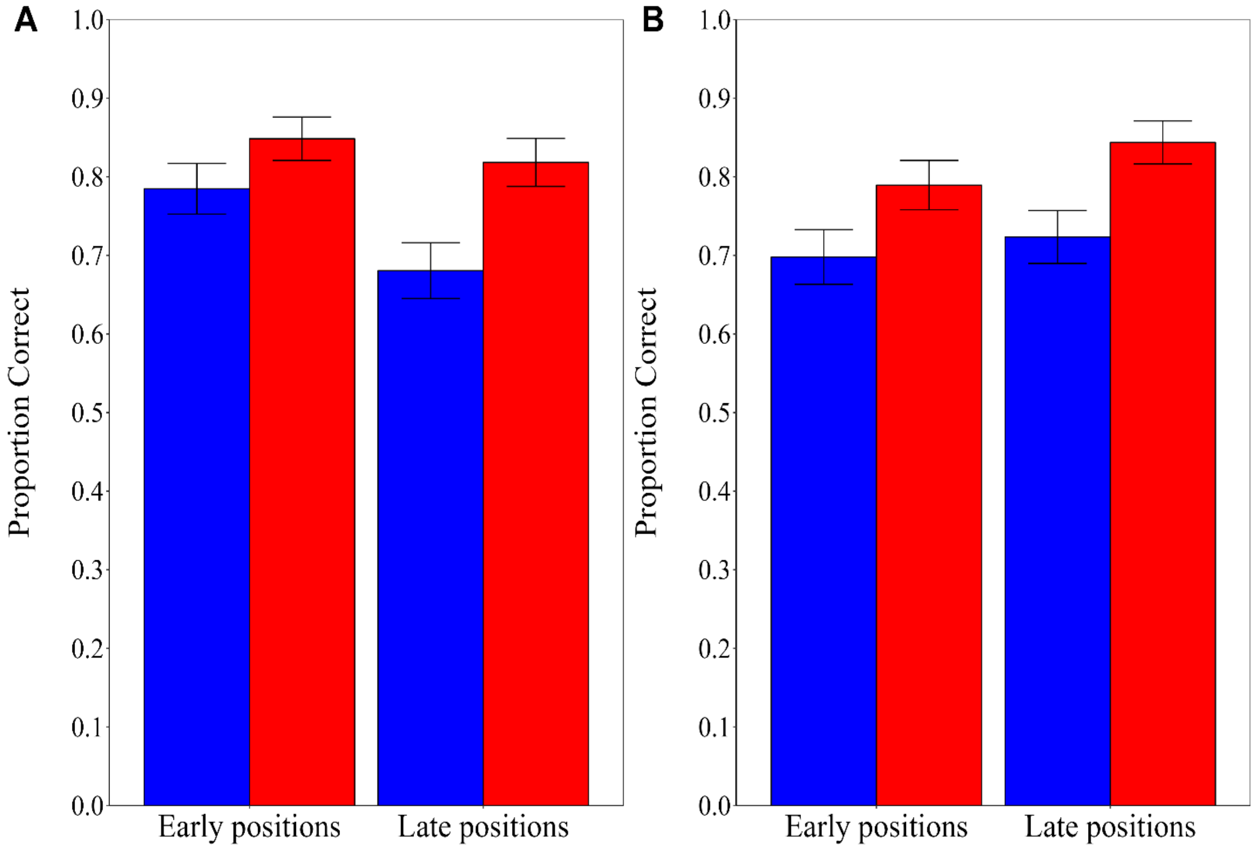
Mean Item Recognition Performance in Experiment 1 as a function of Production Modality, Presentation Order at Test, and Serial Positions



Note. Panels A and B = Direct presentation order. Panels C and D = Random presentation order. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 3

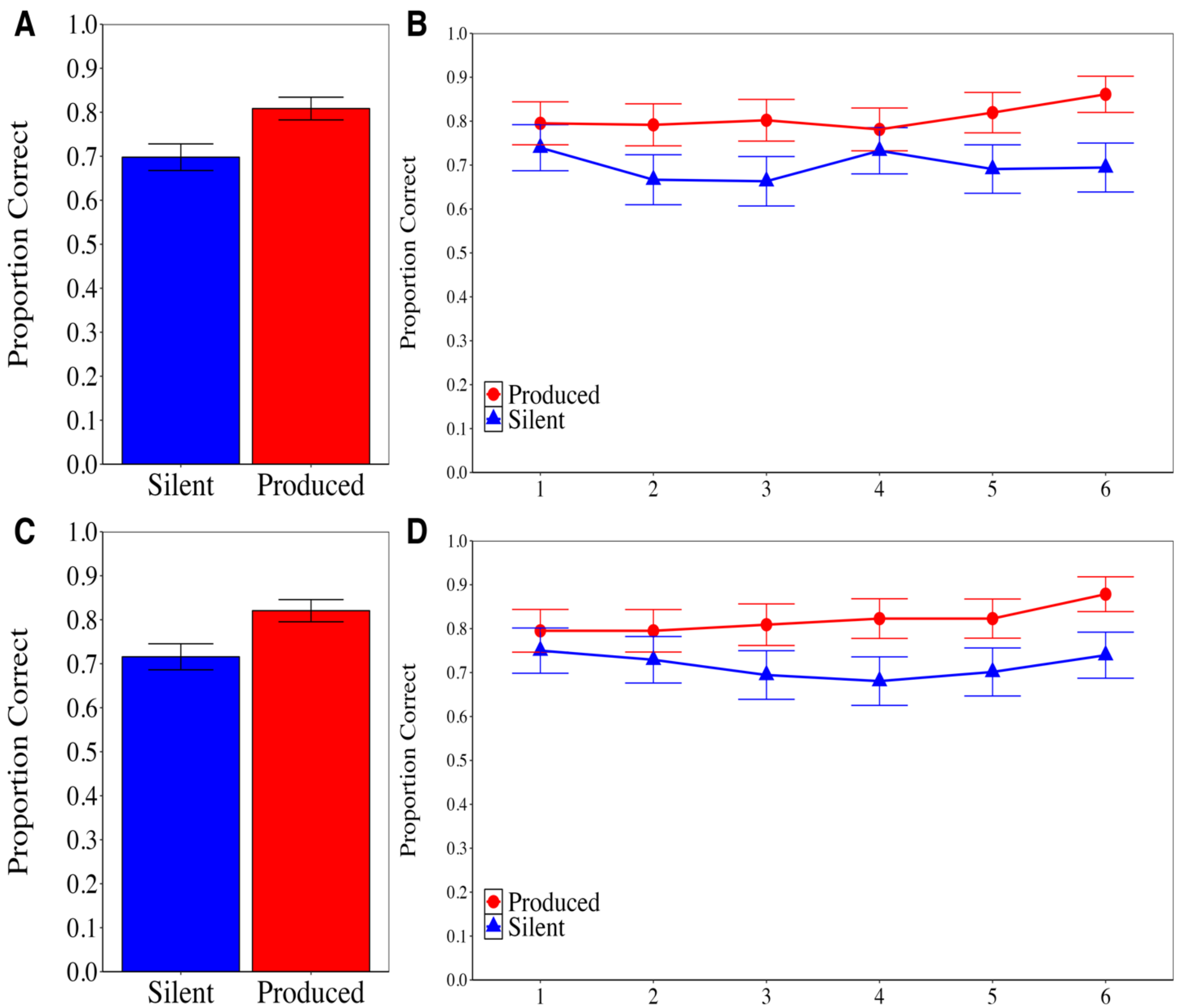
Mean Item Recognition Performance in Experiment 1 as a function of Production Modality, Presentation Order at Test, and Serial Position Block



Note. Red = Produced, Blue = Read silently. Panel A = Direct presentation order. Panel B = Random presentation order. Early positions and Late positions represent the first and last three serial positions. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 4

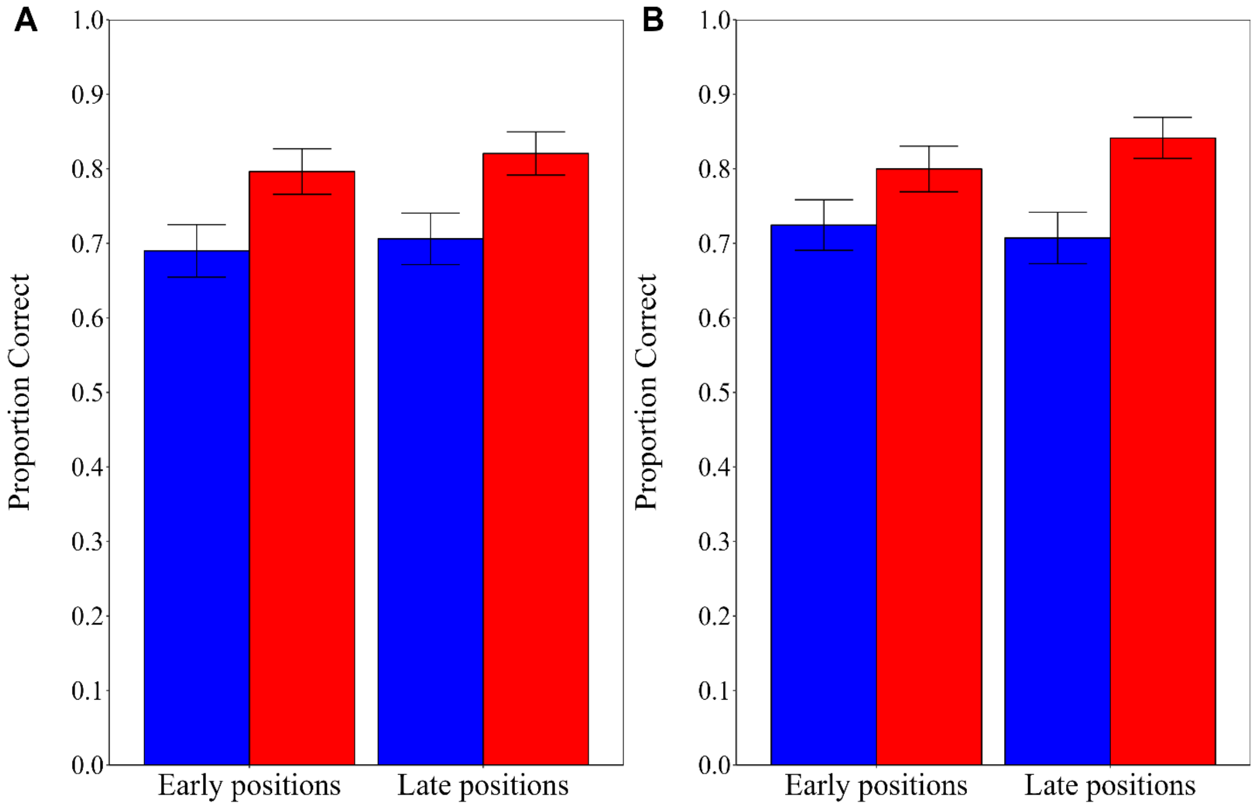
Mean Item Recognition Performance in Experiment 2 as a function of Production Modality, Presentation Order at Test, and Serial Positions



Note. Panels A and B = Direct presentation order. Panels C and D = Random presentation order. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 5

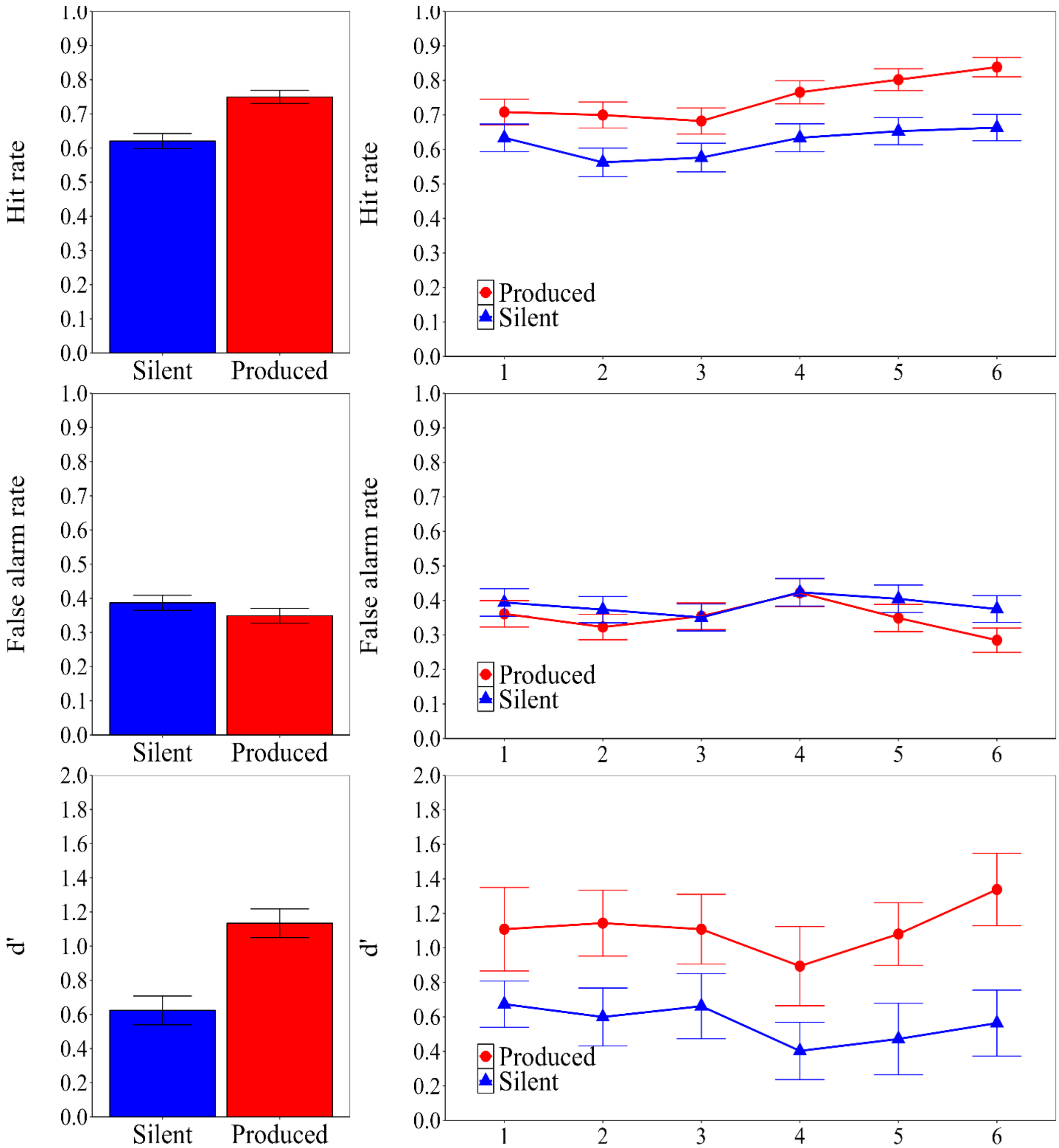
Mean Item Recognition Performance in Experiment 2 as a function of Production Modality, Presentation Order at Test, and Serial Position Block



Note. Red = Produced, Blue = Read silently. Panel A = Direct presentation order. Panel B = Random presentation order. Early positions and Late positions represent the first and last three serial positions. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 6

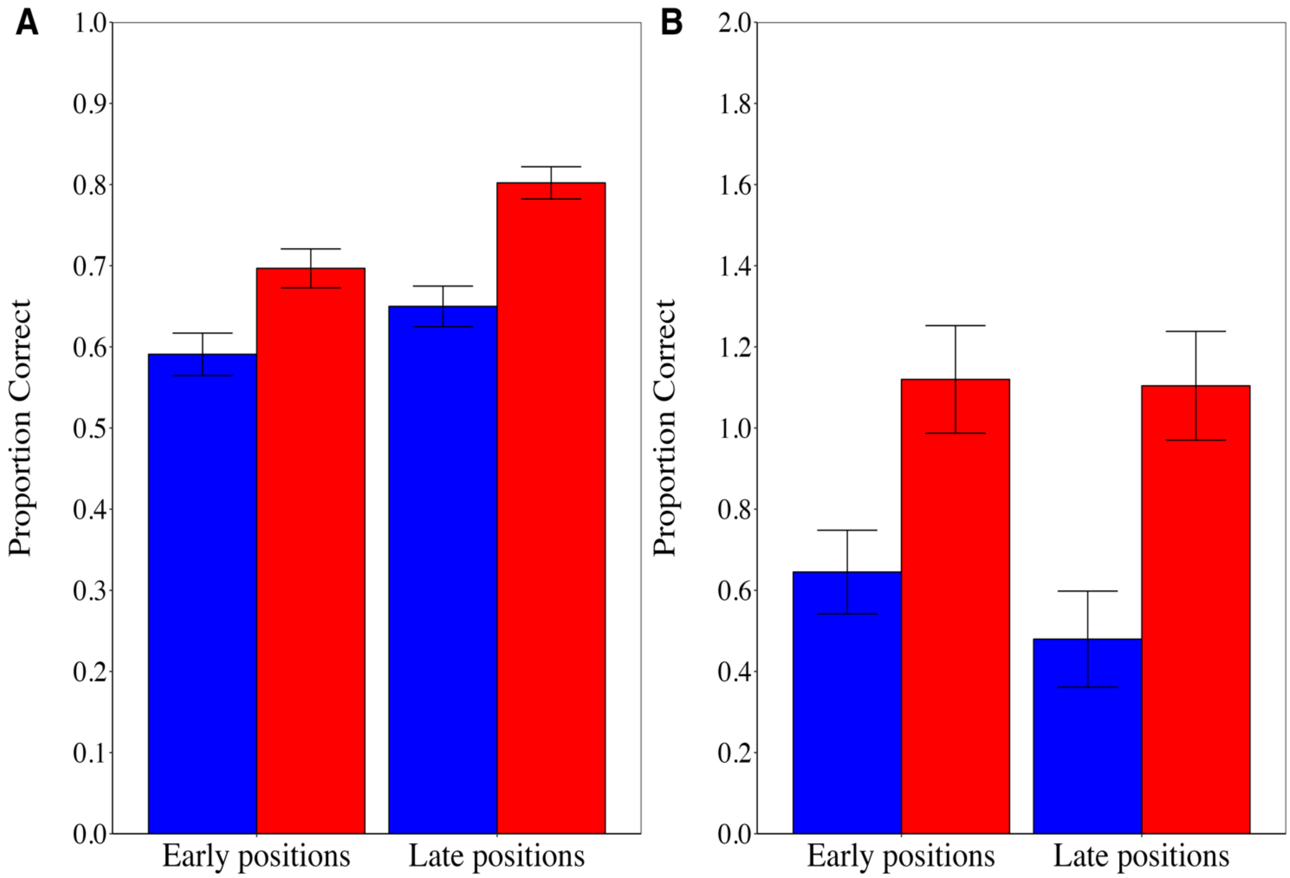
Mean Hit Rate, False Alarm Rate and d' in Experiment 3 as a Function of Production Modality and Serial Position



Note. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 7

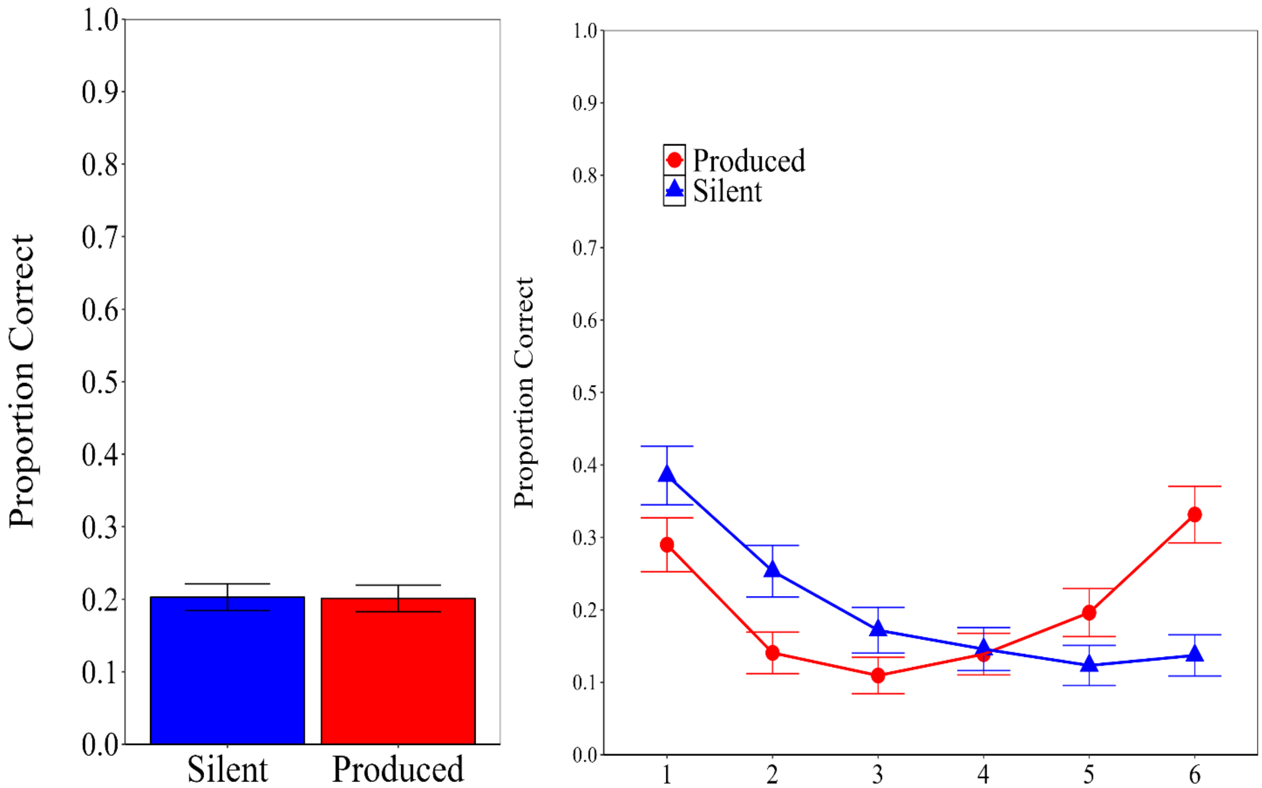
Mean Item Recognition Performance (Hits and d') in Experiment 3 as a function of Production Modality and Serial Position Block



Note. Red = Produced, Blue = Read silently. Panel A = Hits. Panel B = d'. Early positions and Late positions represent the first and last three serial positions. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 8

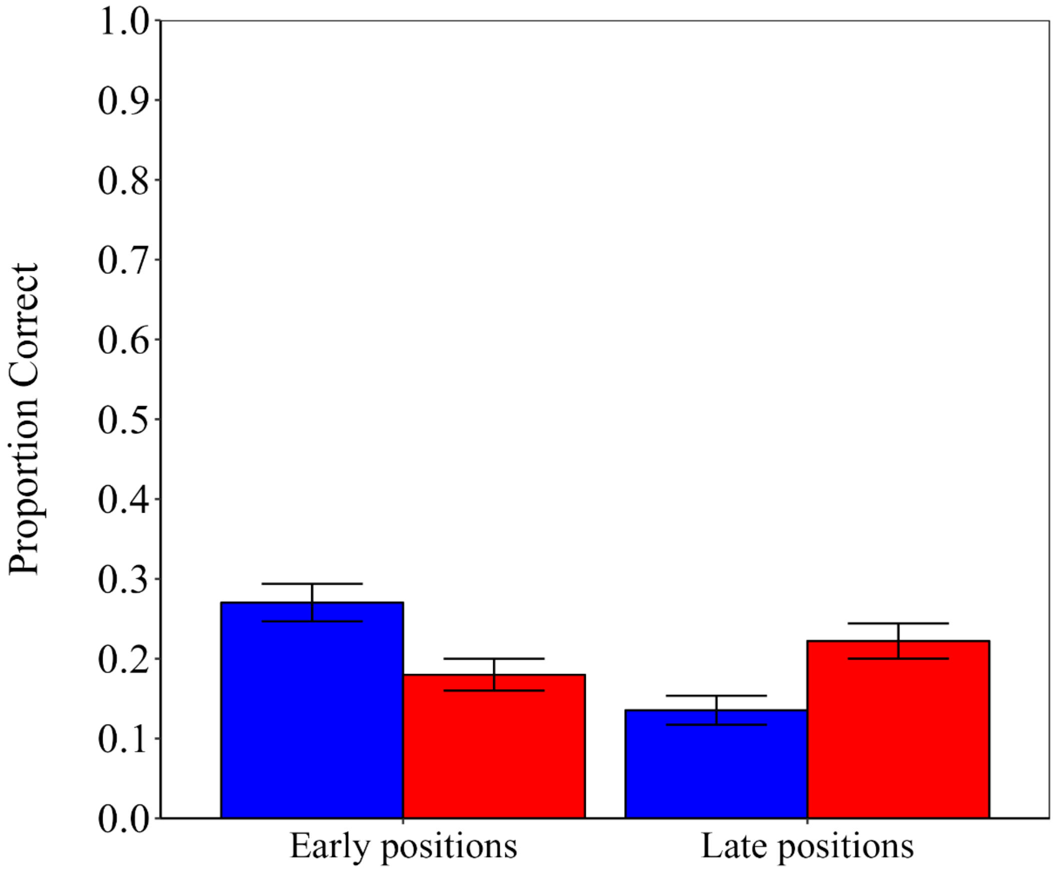
Mean Immediate Serial Recall Performance in Experiment 4 as a function of Production Modality and Serial Positions



Note. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 9

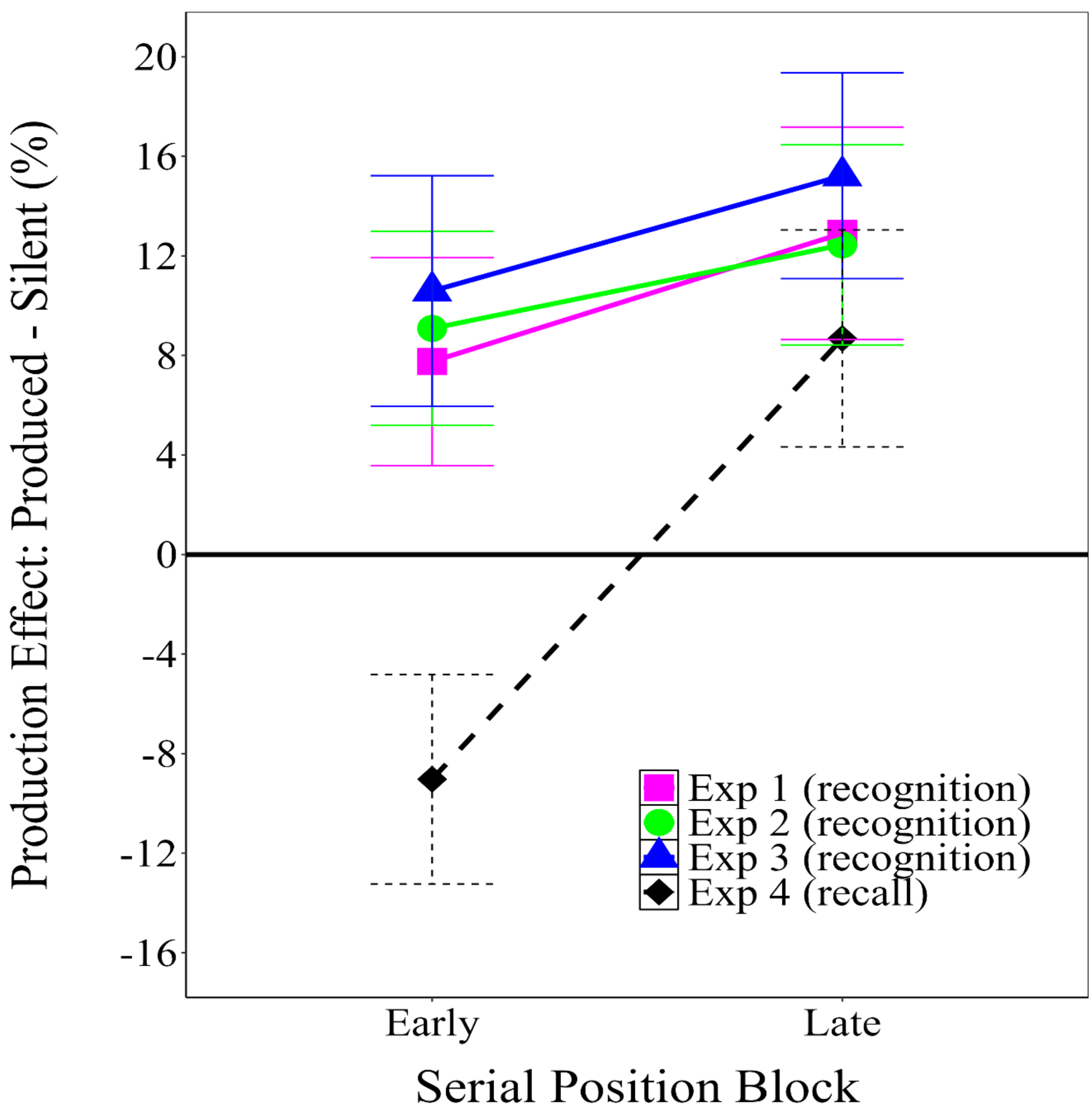
Mean Immediate Serial Recall Performance in Experiment 4 as a function of Production Modality and Serial Position Block



Note. Red = Produced, Blue = Read silently. Early positions and Late positions represent the first and last three serial positions. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure.

Figure 10

Production Effect (Produced – Silent) in All Experiments as a Function of Serial Position Block



Note. Error bars represent 95% confidence intervals computed according to Morey's (2008) procedure. Early positions and Late positions represent the first and last three serial positions. In Experiments 1 and 2, direct and random presentation orders were combined. In Experiment 3, the production effect (produced – silent) was computed based on hit rates.