

Article

Not peer-reviewed version

Perceptual (Static) Active Inference Approach to the Superior Production Effect of Speaking over Writing: An Experiment and Computational Model Report

[Roberto Limongi](#)^{*}, [Oluwagbemisola Oguntoye](#), [Angelica Silva](#)

Posted Date: 16 December 2025

doi: 10.20944/preprints202512.1253.v1

Keywords: production effect; language and memory; memory retrieval



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Perceptual (Static) Active Inference Approach to the Superior Production Effect of Speaking over Writing: An Experiment and Computational Model Report

Roberto Limongi *, Oluwagbemisola Oguntoye and Angelica Silva

Brandon University, Canada

* Correspondence: limongir@brandonu.ca

Abstract

This paper reports a cognitive psychology experiment and a Markov decision process (MDP) model of the production effect—higher memory retrieval that follows speaking aloud or writing/typing words, as opposed to lower memory retrieval when words are read silently. Current models of the production effect draw on the global-matching framework of memory. We identify four limitations of these models and present a MDP model (a perceptual active inference model) to causally explain a superior production effect of speaking over writing. University students performed a word-production task comprising speaking and writing conditions, followed by a memory test. The results showed main effects of condition on accuracy and response times. The MDP model indicated higher sensory precision during memory retrieval in the speaking condition than in the writing condition. Through Bayesian model selection, we evaluated whether the MDP model, as a mechanistic active-inference model, provided higher construct validity than a descriptive linear model (fit via Variational Laplace). The MDP model outperformed the linear model, suggesting that production modalities are hidden states that cause the visual sensory observation of words that had been linguistically produced. Crucially, the MDP model explains both group effects and individual variability, confirming the reliability paradox of statistical models.

Keywords: production effect; language and memory; memory retrieval

1. Introduction

Repeating someone's name aloud right after you're introduced (e.g., "Nice to meet you, Alex!") is a common social interaction strategy in some cultures, and it serves several overlapping purposes. First, it signals attention and respect, helping to establish rapport quickly. It also confirms that you heard the name correctly and provides a smooth transition into the conversation.

Saying the name aloud reinforces memory as well. Names are particularly easy to forget because they're arbitrary and carry little inherent meaning but saying them aloud engages multiple contextual features at once—e.g., auditory stimulation, articulatory actions, and social bonds—making it far more likely that you'll remember the person later. However, what might otherwise be considered a simple manifestation of a cultural or personal preference represents a fundamental cognitive phenomenon of language production and memory that has been scarcely investigated: the production effect.

1.1. The Production Effect: Definition and Experimental Evidence

Visually recognizing words stored in episodic memory—for example, recognizing at test time ($t = 2$) whether a word was visually presented at study time ($t = 1$)—is facilitated when the target word is encoded along with contextual features during the study phase. The cognitive psychology of memory, and especially associative cued-retrieval theories, refer to this phenomenon as the encoding specificity principle [1,2].

In the area of language and memory, higher accuracy in word recognition occurs when participants previously speak a word aloud rather than silently rehearse it. The language and memory literature refers to this phenomenon as the production effect [3]. The production effect occurs not only in the context of spoken words but also in the context of written words [4]. For example, partially typing a word during the study phase negatively affects the production effect at test time [5]. However, the strength of the production effect differs across production modalities. Specifically, even when a word is entirely typed during the production phase, the production effect is weaker than when the word is entirely spoken, constituting a superior production effect of speaking over writing. From an evolutionary perspective, this superior effect likely reflects the earlier emergence of spoken language compared to the later emergence of written language [6].

The distinctiveness account to memory encoding [7,8] applied to the production effect [3,9] indicates that, for example, the motoric act or the auditory perception of the spoken word (and analogously the tactile perception of key presses) represent distinct item-specific contextual features (CF) that the participant encodes along with the target item (I) at study time, constituting joint memory traces (CF, I). This implies that the probability of item recognition (IR) given (CF, I) at test time is greater than the probability of IR given the encoding of I only: $p(\text{IR} | \text{I, CF}) > p(\text{IR} | \text{I})$. Therefore, the distinctiveness account to the production effect positions the production effect as a special case of the encoding specificity principle.

1.2. Towards a Causal Explanation of the Production Effect

The production effect literature either implicitly or explicitly states a causal relationship between contextual features (CF) encoded at study time and item recognition (IR) at test time. For example, MacLeod, Gopie, Hourihan, Neary and MacLeod, Gopie, Hourihan, Neary and Ozubko [3] and Forrin, MacLeod and Ozubko [4] consistently use causal discourse markers [10], such as “produces” and “enhances,” to refer to the production effect. Conforming to this causal discourse, research on the production effect has steadily moved from experimental interventions described through linear models (e.g., t tests, ANOVA, and regression) to explanatory computational models. The first steps in this direction have focused on extensions of models originally formulated within the global-matching framework of memory retrieval [11], in which the encoding specificity principle is a special case.

The global-matching framework refers to a class of computational models of recognition memory that assume that a retrieval cue (e.g., a target word as a concrete I) is compared simultaneously to all stored memory traces and that recognition decisions depend on the aggregate (global) similarity between the cue and the entire memory system. In the special case of the production effect, producing a word aloud creates a memory trace with a greater number of contextual features (CF)—typically operationalized as increased noise-free features or higher encoding strength—compared to silently reading it. During recognition, the probe word (i.e., I) activates all stored traces, and items encoded aloud produce a larger and more salient match signal because their traces are more distinctive.

Drawing on the global-matching framework, the MINERVA 2 model [12] and the REM model [13] have been applied to explain the production effect [14,15]. In MINERVA 2, the activation of stored traces is implemented as a more salient and greater echo intensity due to the presence of more unique features in aloud-encoded traces, whereas in REM the aloud condition generates higher-probability feature matches and lower confusion probabilities in the likelihood ratio driving recognition decisions. Consequently, in both models, production-enhanced traces are more discriminable from lures, leading to higher recognition accuracy. These implementations show that the production effect can be modeled as a distinctiveness-driven increase in the mnemonic signal produced by aloud-encoded items within an associative global-matching architecture. While the global-matching framework has provided an explanatory account of the production effect, relevant computational models such as MINERVA and REM have four main limitations.

First, both schemes are predictive models of associative memory—mapping stimuli to responses—rather than generative causal models comprising latent causal variables (i.e., the unobservable cognitive processes that cause memory retrieval). More specifically, they contain assumptions about how memory traces correlate with each other, but they do not establish cause-effect relations in the way generative models do. It is worth noting, however, that their assumptions are congruent with the traditional notion of memory retrieval and recognition in the cognitive psychology literature [16–19].

Second, the application and interpretability of these models are limited to the memory domain rather than extending to general cognition and brain function.

Third, when applied to the production effect, both models focus only on response accuracy. Although most production-effect research has relied on response accuracy (e.g., proportion of correct probes) as the most important response variable of interest, a few experimental studies have recently incorporated recognition time (indexed by response time) as a secondary dependent measure. For example, recognition time is shorter in self-produced speaking conditions than when production is performed by another participant and faster when compared with silent study [15]. As in any memory-related cognitive phenomenon, accuracy and response times are interdependent variables, and we cannot understand any generative cognitive process of behavior without considering the time these hypothetical processes take [20]. Therefore, relevant computational models are more informative when they account for this interdependence (e.g., the drift diffusion model of memory retrieval) [21].

A fourth limitation of these models refers to the fact that they simulate *in silico* the recognition effect of CF on I but are not applied to empirical data.

Based on the limitations of the global-matching models, a natural next step in the causal agenda of the production effect is to address the question of how CF relevant to the production modality (e.g., speaking and writing words) cause the retrieval of words in a recognition task. We attempt to answer this question experimentally and computationally within the context of perceptual (static) active inference and a Markov decision process (MDP).

1.3. The Current Work

Within the framework of perceptual active inference (cf. predictive coding for perception) [22,23], formalized as a Markov decision process (MDP), we cast cue-dependent retrieval as inference over hidden states that encode contextual episodic features—here, whether a word was previously encoded in a spoken or written context—given the word as the sensory observation. Driven by the task instruction “respond whether you spoke or wrote the word,” the agent embodies a generative model of the causes of its sensory input. This generative model comprises hidden states corresponding to CF (i.e., speaking and writing production modalities) and outcome modalities encoding the observed visual words and the self-observation of responses at test time. Based on these observations, the agent updates its beliefs about the latent CF by inverting the generative model, selecting the CF state that minimizes variational free energy. From this perspective, what is treated as cue-dependent retrieval in the global-matching framework corresponds, in our causal generative model, to belief selection that maximizes model evidence by favoring the spoken or written CF state whose predicted sensory consequences best account for the observed word and the self-observed response. This formulation thus provides an active-inference realization of the encoding specificity principle in terms of state estimation under an MDP.

The current work builds on previous findings showing that recognition accuracy is higher for spoken than for written words [4,6]. Extending these results, we incorporate within a single model a parameter capturing the hypothesis that the stronger production effect observed in the spoken modality is also associated with shorter response times (RTs). We reiterate that influential cognitive models of memory recognition such as the drift diffusion model [21] account for both accuracy and response times within a unified framework, reflecting their interdependence (e.g., the speed-accuracy trade-off). Accordingly, we test the hypothesis that RTs are shorter when recognizing words

encoded in the spoken CF than in the written CF. In essence, by specifying a single perceptual model, we simultaneously test the novel hypothesis that recognition is faster in the speaking condition than in the writing condition and that a causal generative model provides a better explanation of the data than simple statistical linear models.

2. Materials and Methods

2.1. Participants

Under laboratory conditions, 16 undergraduate students (6 biological males and 10 biological females; M age = 22.5, SD = 6.22) from Brandon University were recruited via campus advertisements. Sample size estimation was based on sequential analysis and optional stopping [24,25] (described below). Participation was voluntary, and all individuals provided written informed consent prior to the initiation of the experiment. All participants received a \$50 gift card as compensation for their time. The study was approved by the Brandon University Research Ethics Committee.

Optional Stopping for Sample Size Definition

After collecting data from 10 participants, we fit Bayesian ANCOVA models to the RT and accuracy data, including condition as a fixed effect, participant as a random effect, and word frequency as a covariate. All ANCOVA models were implemented in JASP [26]. We defined a stopping criterion based on $BF_{10} > 10$, indicating strong evidence in support of shorter RTs and higher accuracy in the speaking condition compared with the writing condition. Data collection proceeded one participant at a time, with BF_{10} recomputed after each addition, until the threshold was reached.

2.2. Stimuli

Our stimuli consisted of an experimental word list of 180 nouns with the highest academic frequency selected from the Strathy Corpus of Canadian English [27]. For each participant, the order of the 180 words was randomized, and the list was split into two sublists (one per condition) of 90 words each. This randomized order was used to assign each word, without replacement, to a language condition (writing or speaking). Nine block lists of 20 words each were then created, with words randomly drawn from the speaking and writing sublists (10 words per condition).

2.3. Procedure

Participants were seated at a computer and wore noise-cancelling headphones throughout the experiment. Each participant completed nine blocks, each consisting of a language production phase followed by a memory phase (Figure 1). After completing each block, participants took a break and were free to decide when to begin the next block by pressing the space bar. The experiment began with four familiarization trials (two per condition). In the production phase of each block, participants were presented with 20 words for 5 s each, displayed either in red ink (speaking condition) or green ink (writing condition). The color of the word cued participants to either speak or type the word, respectively. The color-condition association was counterbalanced across participants, with the mapping reversed for half of the sample. Each word presentation was preceded by a "be prepared" slide displayed for 4.5 s. After completing the language production phase, participants performed the memory phase using the same words presented during the production phase. Probe words were displayed in black ink on a white background for 5 s, unless a response was recorded earlier. Feedback was then provided ("correct," "incorrect," or "no response recorded") for 5 s before the next trial began. The task was built using PsychoPy, which recorded accuracy, reaction times, and word production. Accuracy of word production (i.e., whether the participant actually spoke or wrote the target word was confirmed via PsychoPy output (for written registers) and spectrogram analysis (for spoken registers).

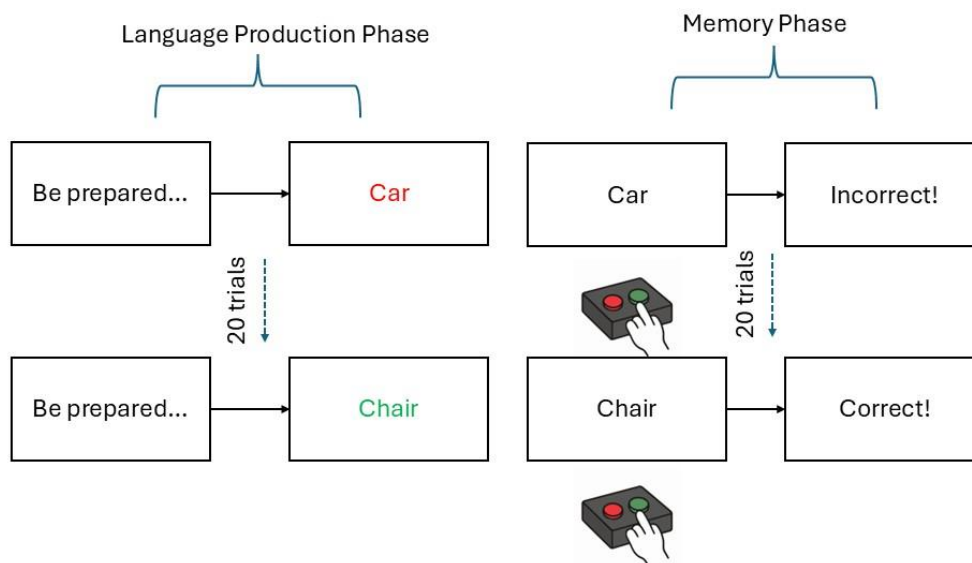


Figure 1. Experimental task block. Participants performed 9 blocks of 20 trials in two phases: language production phase and memory phase. As shown in the production phase, the language modality was cued by the word color (speaking in red, writing in green) counterbalanced across participants. In the memory phase, participants responded by pressing relevant keys whether the word was spoken or written (colored buttons do not represent the actual response keys which were "S" for speaking and "W" for writing).

2.4. MDP Model

We used SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>) and modified scripts provided by Smith, *et al.* [28] and Smith, *et al.* [29] to specify and estimate the parameters of a two-timestep MDP model that mirrored the experimental task. The model included one state factor, two outcome modalities, and one state transition matrix. The state factor comprised three states: a start state, a speaking context, and a writing context (as CF). The first outcome modality comprised three possible observations: no response, spoken word, and written word. The second outcome modality comprised self-observations of correct and incorrect responses. The state transition matrix encoded equal probabilities of transitioning from the start state to either the speaking or writing context. Transitions between language-production contexts were modeled as absorbing states; once the agent transitioned from the start state into a context state, it remained there. Figure 2 illustrates the factor graph and matrices of the MDP model.

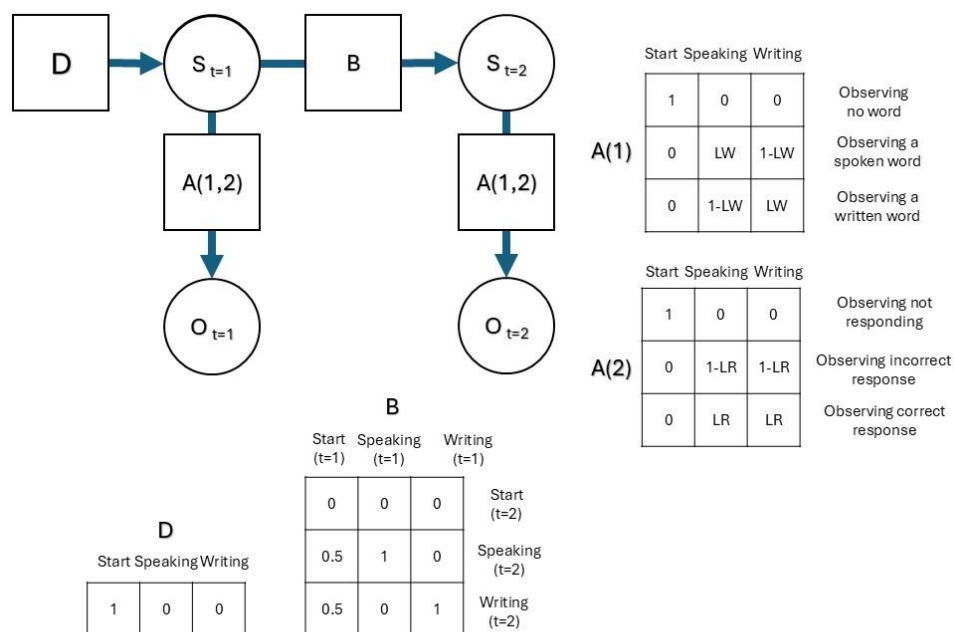


Figure 2. Factor graph of the MDP model. LR (L-Response parameter), LW (L-Word parameter). D matrix (state beliefs at $t=1$), B matrix (state transition probabilities between $t=1$ and $t=2$), A matrices (likelihood matrices of two outcome modalities).

The response model assumed a stochastic action-selection process in which the choice of a language-production modality (i.e., CF) corresponded to sampling from the posterior distribution over the hidden (true) state given the observations (spoken or written word and response accuracy). Based on our hypothesis that the speaking CF causes faster recognition times than the writing CF, the response model followed the rule that, as the posterior probability of being in the speaking-CF state—as the cause of visual word recognition—approached 1 after observing a probe word and self-observing an accurate response, the probability of an early response increased relative to the probability of a late response, and vice versa.

For improved computational tractability and interpretability of the MDP results, we discretized participants' response times (RTs) into two categories: early and late responses. At the participant level and across conditions, we used the median of the RT distribution as the cutoff point. In forced-choice decision-making tasks, RTs typically follow an ex-Gaussian distribution. Discretizing RTs drawn from an ex-Gaussian distribution using the median provides a practical method for converting highly skewed continuous variables into categorical observations. The ex-Gaussian RT distribution combines a Gaussian component with a positively skewed exponential tail. In this type of distribution, the median serves as a robust measure of central tendency that is insensitive to extreme values in the exponential tail, ensuring that the resulting early- versus late-response categories reflect the empirical structure of the data rather than being disproportionately influenced by outliers. Moreover, median-based discretization is particularly convenient in perceptual active-inference models, where RT categories serve as outcome modalities. This approach enhances model interpretability while preserving meaningful information inherent in the original RT data.

3. Results

Behavioral Results

Table 1 summarizes the behavioral results. Mean RT was shorter in the speaking condition than in the writing condition. Similarly, the mean proportion of correct responses was higher in the

speaking condition than in the writing condition. Bayesian ANCOVA models confirmed these differences and ruled out an effect of word frequency (i.e., the model including the covariate received less support, $BF_{10} = 1390$, than the model without the covariate, $BF_{10} = 3899$; Figure 3). A similar Bayesian ANCOVA model confirmed the accuracy difference and again ruled out an effect of word frequency ($BF_{10} = 325$ vs. $BF_{10} = 59$, respectively; Figure 3). We further fit a third ANCOVA model to confirm that the response-time difference between conditions remained after discretization ($BF_{10} = 3948$); the model including word frequency as a covariate was less supported by the data ($BF_{10} = 335$; Figure 3).

Table 1. Descriptive Statistics.

	Speaking (M \pm SD)	Writing (M \pm SD)
Number of Trials	82.50 \pm 8.76	82.94 \pm 8.07
Reaction Time (s)	1.64 \pm 0.38	1.77 \pm 0.40
Proportion Correct	0.88 \pm 0.07	0.81 \pm 0.12
Proportion Incorrect	0.12 \pm 0.07	0.19 \pm 0.12

Note. Values represent subject-level means (M) and standard deviations (SD).

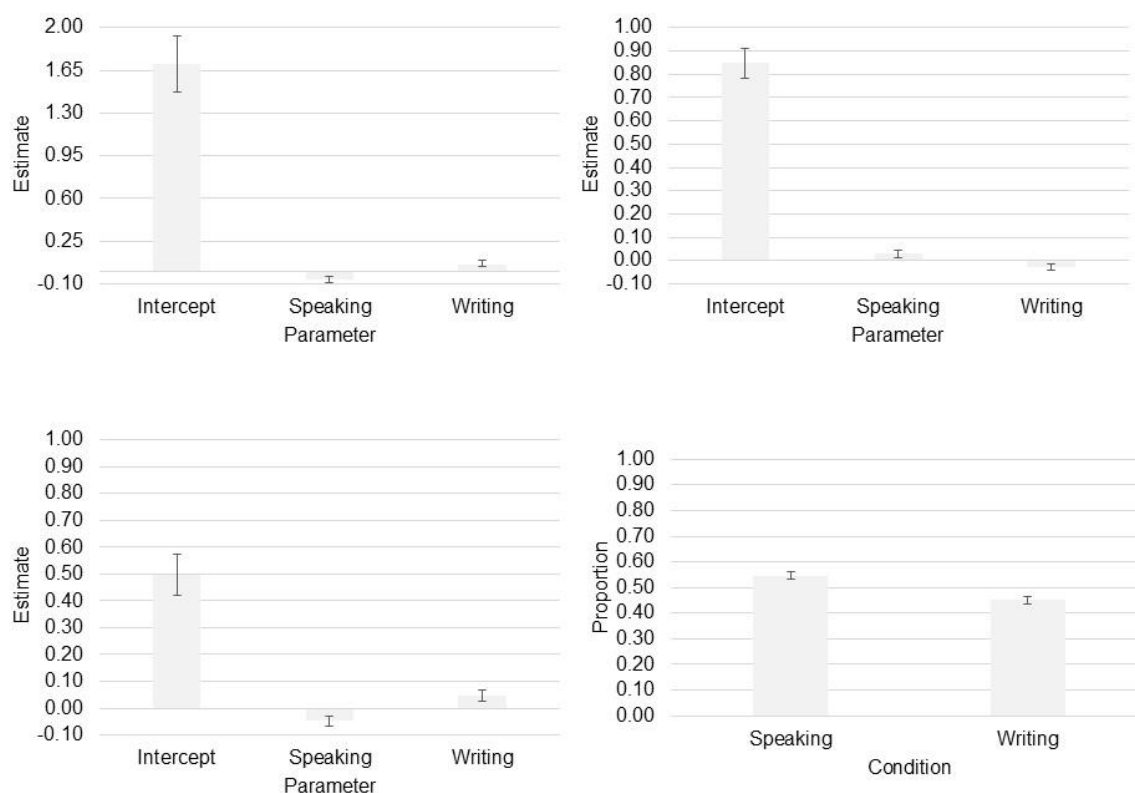


Figure 3. Parameter estimates of Bayesian ANCOVA models. Effect of condition on RT (top left), effect of condition on response (word memory retrieval) accuracy (top right). Difference in response time held after discretization (bottom left). Bottom right panel shows the group-mean proportions of early (discretized) responses in the speaking and writing conditions, mirroring the behavioral pattern of the raw RTs.

4. Computational Model Results, Linear Model Under VL, and BMS

Table 2 shows the parameter estimates of the MDP model. The mean parameter estimate for L-Word was 0.55 (SD = 0.06), whereas the mean parameter estimate for L-Response was 0.54 (SD = 0.04).

Bayesian one-sample t tests (implemented in JASP) confirmed that both parameters differed from 0.5 ($BF_{10} = 9.34$ and $BF_{10} = 84$, respectively, Figure 4).

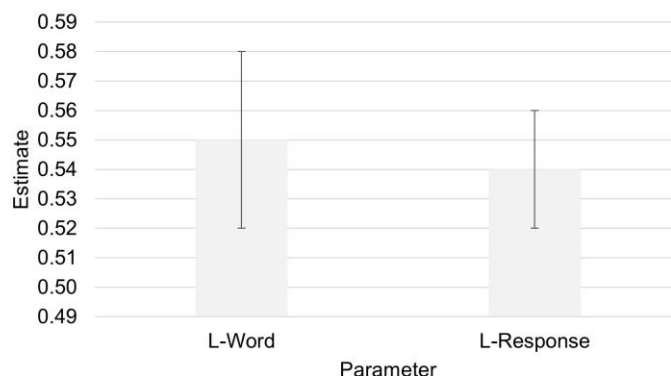


Figure 4. Parameter estimates of the MDP model (group level). Error bars correspond to the 95% credible interval.

Because the MDP embodies the theoretical construct that production modality is a hidden cause of the observations, we assessed the construct validity of this model by comparing it against a simple linear model fit under VL [30] to discretized response times, with condition and accuracy as predictors. Crucially, VL allows for the comparison of models with different likelihood functions using Bayesian model selection. Therefore, we adjudicated between the causal generative model and the linear model based on their free energies. Bayesian model selection supported the MDP, yielding a protected exceedance probability of 0.94 and a Bayesian omnibus risk (BOR) of less than 0.001.

Table 2. Parameter estimates of the MDP model.

Participant	L-Word	L-Response
001	0.48	0.55
002	0.65	0.56
003	0.57	0.60
004	0.56	0.59
005	0.58	0.46
006	0.50	0.50
007	0.56	0.53
008	0.52	0.54
009	0.58	0.57
010	0.59	0.51
011	0.56	0.50
012	0.35	0.57
013	0.53	0.59
014	0.55	0.52
015	0.57	0.52
016	0.58	0.55

5. Discussion

We put forward the idea that the MDP embodies the theoretical construct that production modality is a hidden cause of word memory retrieval. The superior model evidence for the MDP,

relative to a simple statistical description model, provides computational support for the construct validity of this hidden-state formulation. The reported MDP model also provides evidence that speaking words aloud reduces word retrieval time compared to typing words. To our knowledge, this is the first attempt to explain the production effect using computational models specified within the active inference framework, albeit at the level of static perception or predictive coding.

Within the field of computational models of cognition, many authors argue that such models provide a deeper understanding of cognitive phenomena than descriptions based solely on statistical models [31–35]. However, direct comparisons between computational models and statistical models remain scarce. Crucially, this study provides the first direct model comparison of the production effect and, to our knowledge, the first comparison between linear models estimated using VL and MDP evaluated via Bayesian model selection.

Crucially, the MDP model confirms the reliability paradox of statistical models, whereby consistent group effects [...] do not accurately reflect individual differences [33]. The 95% credible intervals for both the L-Word and L-Response parameters indicate that the group-level parameters exceed $p = 0.5$. Although the effect sizes appear small when expressed in terms of probabilities, they correspond to moderate-to-large and large effects in standardized terms. For context, consider the equivalent frequentist null-hypothesis tests at the group level: for L-Word, $t(15) = 4.1$, $p < 0.001$, CI [0.02, 0.06]; and for L-Response, $t(15) = 2.8$, $p = 0.012$, CI [0.012, 0.081]. The corresponding Cohen's d values are 0.71 for the L-Word parameter (medium-to-large effect size) and 1.03 for the L-Response parameter (large effect size). The robust group-level effect sizes do not mask individual variability. Specifically, we identified two participants who showed higher precision in the writing contextual state and one participant who showed higher precision in the self-observation of incorrect responses (Table 2). These individual differences raise at least two questions regarding the production effect.

First, differences in the L-Response parameter indicate inter-individual variability in the precision of self-monitoring of response outcomes. Most participants showed relatively reliable inference about their own performance, whereas two participants exhibited noisier self-evaluation. Such results provide a basis for characterizing individual cognitive phenotypes. One possible interpretation is that these participants are more prone to detecting their errors than their successes during language production, a distinction that, as elaborated upon below, is particularly relevant in the context of language learning.

In the second language acquisition (SLA) literature, learners differ systematically in their sensitivity to negative evidence, with some individuals exhibiting an error-oriented self-monitoring style in which deviations from target forms are more salient than successful productions. This phenomenon has been discussed under labels such as heightened responsiveness to negative evidence, over-monitoring, and differential noticing of errors, particularly in research on focus on form and feedback processing [36–40].

From the perspective of static perception or predictive coding, however, such individual differences do not need to be framed as strategic choices or motivational dispositions, but can instead be understood as differences in the precision-weighting of self-generated prediction errors during perceptual inference [41,42]. In a perceptual Markov decision process, self-evaluative outcomes—such as perceived correctness or incorrectness of a response—are treated as observations generated by latent performance states. Learners who are more error-oriented can thus be modeled as assigning higher precision to error-related self-observations, such that prediction errors signaling incorrect performance exert a stronger influence on posterior beliefs than signals of success. Conversely, learners who are less focused on errors effectively down-weight these prediction errors, treating them as noisier or less informative. This computational account reframes SLA notions of error orientation and sensitivity to negative evidence not as a bias toward negative outcomes per se, but as principled differences in perceptual precision within a Bayesian generative model, thereby providing a mechanistic explanation for individual differences in self-monitoring during second language learning.

Second, most participants were more likely to infer the production modality of spoken words than that of written words, as indicated by the group-level L-Word parameter, leading to faster recognition times. Why, then, did participants take longer to recognize written words? As noted in the Introduction, evolutionary and cultural factors may help explain the superior production effect of spoken words, for which fast and brief utterances are common. Written language, by contrast, was not culturally developed to facilitate rapid communication. Writing is inherently time-consuming because it is closely tied to the exploration and generation of knowledge. In other words, spoken language tends to be less elaborate and typically shorter in informational content and lexical density than written texts such as textbooks and academic prose [43].

Furthermore, although retrieving individual words is a prerequisite for language production, it is not necessarily indicative of knowledge generation, which is a core function of written language. Writing is therefore a slower process, a property that is congruent with active inference accounts emphasizing exploratory behavior. A natural follow-up question is whether producing more complex registers (e.g., sentences combining multiple words) in spoken language confers advantages over writing [44] in memory retrieval of individual words. That is, language users may be more accustomed to producing written registers to communicate extended bodies of knowledge or to generate new knowledge. While this explanation offers a computational perspective on evolutionary and cultural differences across language modalities, it also opens the possibility that slower recognition times in the writing condition constitute a computational advantage from the perspective of the exploration–exploitation dilemma. In this view, the few participants who showed faster recognition times in the writing modality may have resolved this dilemma in line with their individual preferences. This point highlights one of several limitations of the current model and experimental task, which speaks to future improvements.

First, as a static perceptual model, it does not exploit the full computational and theoretical capabilities of the active inference framework, particularly the possibility of a speaking–writing trade-off between exploration and exploitation. Addressing this trade-off would require introducing a third timestep in the model, in which participants could choose not to respond at $t = 2$ and respond at $t = 3$ (corresponding to late responses), or respond at $t = 2$ (corresponding to early responses). Incorporating these possible actions would also require specifying at least two shallow policies (i.e., wait and respond). Future work should explore this extended active inference scheme.

Second, we discretized response times (RTs). However, the memory research literature has long emphasized the value of analyzing the full RT distribution. This is feasible within the active inference framework. Future work could, for example, directly compare a model using discretized RTs with a continuous-state model.

Third, the present study compared a perceptual model with a linear model, demonstrating a better account of the data and stronger causal construct validity. However, this model should also be directly compared with established memory models of the production effect (e.g., MINERVA and REM).

Finally, from an experimental task design perspective, this study focused on a specific hypothesis concerning the superior production effect of speaking over writing (i.e., a relative production effect). However, it did not address the absolute production effect of either modality by including a control condition (e.g., silently reading words).

Conclusions

While previous research has provided experimental evidence for the production effect and initial computational accounts of its dynamics, the present work advances the understanding of this language and memory phenomenon by proposing a generative model of the causes underlying faster retrieval times for spoken compared with written words. The study also provides direct evidence for the superior explanatory power of active inference models relative to simpler statistical models. Notably, this comparison represents the first use of a linear model estimated under the VL scheme for this purpose. Importantly, the proposed model simultaneously identifies group-level parameters

that capture the general dynamics of the production effect and reveals individual differences in parameter estimates, with potential implications for language learning and knowledge generation through writing.

Author Contributions: Conceptualization, R.L.; methodology, R.L, O.O, and A.S.; writing—original draft preparation, R.L and O.O.; writing—review and editing, R.L, O.O, and A.S.; supervision, R.L.; funding acquisition, R.L. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: “This research was funded by the Social Sciences and Humanities Research Council, Insight Development Grant 430-2024-00727”.

Institutional Review Board Statement: “The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Brandon University (protocol code 23412 October 2, 2024).

Data Availability Statement: Data and scripts are available at <https://people.brandonu.ca/limongir/home/the-writing-brain-laboratory/>

Conflicts of Interest: “The authors declare no conflicts of interest.”

References

1. Tulving, E.; Thomson, D.M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* **1973**, *80*, 352-373, doi:10.1037/h0020071.
2. Fisher, R.P.; Craik, F.I.M. Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory* **1977**, *3*, 701-711, doi:10.1037/0278-7393.3.6.701.
3. MacLeod, C.M.; Gopie, N.; Hourihan, K.L.; Neary, K.R.; Ozubko, J.D. The production effect: delineation of a phenomenon. *J Exp Psychol Learn Mem Cogn* **2010**, *36*, 671-685, doi:10.1037/a0018785.
4. Forrin, N.D.; MacLeod, C.M.; Ozubko, J.D. Widening the boundaries of the production effect. *Memory & cognition* **2012**, *40*, 1046-1055, doi:10.3758/s13421-012-0210-8.
5. Kelly, M.O.; Ensor, T.M.; MacLeod, C.M.; Risko, E.F. The prod eff: Partially producing items moderates the production effect. *Psychon Bull Rev* **2024**, *31*, 373-379, doi:10.3758/s13423-023-02360-9.
6. MacLeod, C.M.; Bodner, G.E. The Production Effect in Memory. *Current Directions in Psychological Science* **2017**, *26*, 390-395.
7. Reed Hunt, R. The Concept of Distinctiveness in Memory Research. In *Distinctiveness and Memory*, Hunt, R.R., Worthen, J.B., Eds.; Oxford University Press: 2006; p. 0.
8. Hunt, R.R. Precision in Memory Through Distinctive Processing. *Current Directions in Psychological Science* **2013**, *22*, 10-15, doi:10.1177/0963721412463228.
9. Icht, M.; Mama, Y.; Algom, D. The production effect in memory: multiple species of distinctiveness. *Front Psychol* **2014**, *5*, 886, doi:10.3389/fpsyg.2014.00886.
10. Chun, C.; Lee, S.; Seo, J.; Lim, H. CReTIHC: Designing Causal Reasoning Tasks about Temporal Interventions and Hallucinated Confoundings. Singapore, December, 2023; pp. 10334-10343.
11. Osth, A.F.; Dennis, S. Global Matching Models of Recognition Memory. *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications* **2024**, *0*, doi:10.1093/oxfordhb/9780190917982.013.31.
12. Hintzman, D.L. MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers* **1984**, *16*, 96-101, doi:10.3758/BF03202365.
13. Shiffrin, R.M.; Steyvers, M. A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review* **1997**, *4*, 145-166, doi:10.3758/BF03209391.
14. Jamieson, R.K.; Mewhort, D.J.K.; Hockley, W.E. A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale* **2016**, *70*, 154-164, doi:10.1037/cep0000081.
15. Kelly, M.O.; Ensor, T.M.; Lu, X.; MacLeod, C.M.; Risko, E.F. Reducing retrieval time modulates the production effect: Empirical evidence and computational accounts. *Journal of Memory and Language* **2022**, *123*, 104299, doi:https://doi.org/10.1016/j.jml.2021.104299.

16. Tulving, E. Elements of episodic memory. 1983.
17. Yonelinas, A.P.; Ramey, M.M.; Riddell, C. Recognition Memory: The Role of Recollection and Familiarity. In *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*, Kahana, M.J., Wagner, A.D., Eds.; Oxford University Press: 2024; p. 0.
18. Mandler, G. Recognizing: The judgment of previous occurrence. *Psychological Review* **1980**, *87*, 252-271, doi:10.1037/0033-295X.87.3.252.
19. Anderson, M.C. Retrieval. In *Memory*, Baddeley, A., Eysenck, M.W., Anderson, M.C., Eds.; Routledge: 2025; pp. 299 - 342.
20. Luce, R.D. *Response Times*; 1991.
21. Ratcliff, R. A theory of memory retrieval. *Psychological Review* **1978**, *85*, 59-108, doi:10.1037/0033-295X.85.2.59.
22. Sprevak, M.; Smith, R. An Introduction to Predictive Processing Models of Perception and Decision-Making. *Top Cogn Sci* **2023**, doi:10.1111/tops.12704.
23. Hodson, R.; Mehta, M.; Smith, R. The empirical status of predictive coding and active inference. *Neuroscience and biobehavioral reviews* **2024**, *157*, 105473, doi:10.1016/j.neubiorev.2023.105473.
24. Rouder, J.N. Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review* **2014**, *21*, 301-308, doi:10.3758/s13423-014-0595-4.
25. Tendeiro, J.N.; Kiers, H.A.L.; van Ravenzwaaij, D. Worked-out examples of the adequacy of Bayesian optional stopping. *Psychonomic Bulletin & Review* **2022**, *29*, 70-87, doi:10.3758/s13423-021-01962-5.
26. Team, J. *JASP (Version 0.18) [Computer software]*; <https://jasp-stats.org/>: 2024.
27. Unit, S.L. *Strathy Corpus of Canadian English*; 2024.
28. Smith, R.; Friston, K.J.; Whyte, C.J. A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology* **2022**, *107*, 102632, doi:10.1016/j.jmp.2021.102632.
29. Smith, R.; Kuplicki, R.; Feinstein, J.; Forthman, K.L.; Stewart, J.L.; Paulus, M.P.; Tulsa, i.; Khalsa, S.S. A Bayesian computational model reveals a failure to adapt interoceptive precision estimates across depression, anxiety, eating, and substance use disorders. *PLOS Computational Biology* **2020**, *16*, e1008484, doi:10.1371/journal.pcbi.1008484.
30. Zeidman, P.; Friston, K.; Parr, T. A primer on Variational Laplace (VL). *Neuroimage* **2023**, *279*, 120310, doi:<https://doi.org/10.1016/j.neuroimage.2023.120310>.
31. Smaldino, P.E. How to Translate a Verbal Theory Into a Formal Model. *Social Psychology* **2020**, *51*, 207-218, doi:10.1027/1864-9335/a000425.
32. Jarecki, J.B.; Tan, J.H.; Jenny, M.A. A framework for building cognitive process models. *Psychon Bull Rev* **2020**, *27*, 1218-1229, doi:10.3758/s13423-020-01747-2.
33. Haines, N.; Kvam, P.D.; Irving, L.; Smith, C.T.; Beauchaine, T.P.; Pitt, M.A.; Ahn, W.-Y.; Turner, B.M. A tutorial on using generative models to advance psychological science: Lessons from the reliability paradox. *Psychological Methods* **2025**, No Pagination Specified-No Pagination Specified, doi:10.1037/met0000674.
34. Heathcote, A.; Brown, S.D.; Wagenmakers, E.-J. An Introduction to Good Practices in Cognitive Modeling. In *An Introduction to Model-Based Cognitive Neuroscience*, Forstmann, B.U., Wagenmakers, E.-J., Eds.; Springer New York: New York, NY, 2015; pp. 25-48.
35. Turner, B.M.; Forstmann, B.U.; Love, B.C.; Palmeri, T.J.; Van Maanen, L. Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology* **2017**, *76*, 65-79, doi:<https://doi.org/10.1016/j.jmp.2016.01.001>.
36. Schmidt, R. The Role of Consciousness in Second Language Learning¹. *Applied Linguistics* **1990**, *11*, 129-158, doi:10.1093/applin/11.2.129.
37. Schmidt, R. Attention. In *Cognition and Second Language Instruction*, Robinson, P., Ed.; Cambridge Applied Linguistics; Cambridge University Press: Cambridge, 2001; pp. 3-32.
38. Long, M. The role of the linguistic environment in second language acquisition. *Handbook of second language acquisition* **1996**.
39. Gass, S.M.; Gass, S.M. *Input, Interaction, and the Second Language Learner*; Routledge, 1997.
40. DeKeyser, R. Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. *Focus on form in classroom second language acquisition* **1998**, *28*, 42-63.

41. Friston, K. The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences* **2009**, *13*, 293-301, doi:<https://doi.org/10.1016/j.tics.2009.04.005>.
42. Feldman, H.; Friston, K.J. Attention, uncertainty and free-energy. *Front Hum Neurosci* **2010**, *4*, 1-23, doi:10.3389/fnhum.2010.00215.
43. Biber, D. *Variation across speech and writing*; Cambridge University Press: Cambridge, UK, 1988.
44. Skelton, A.B.; De Vries, L.A.; Silva, A.M.; Limongi, R. Writing through Sympathetic Arousal: A Perceptual Active Inference Model on the Effects of Sympathetic Arousal Interoception on Written and Spoken Grammatical Complexity. In Proceedings of the Proceedings of the 6th International Workshop on Active Inference, 2025, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.