



HAL
open science

Bayesian modeling of lexical knowledge acquisition in BRAID, a visual word recognition model

Emilie Ginestet, Sylviane Valdois, Julien Diard

► **To cite this version:**

Emilie Ginestet, Sylviane Valdois, Julien Diard. Bayesian modeling of lexical knowledge acquisition in BRAID, a visual word recognition model. Conference of the Society of Scientific Studies of Reading (SSSR), Jul 2018, Brighton, United Kingdom. hal-02004280

HAL Id: hal-02004280

<https://hal.science/hal-02004280>

Submitted on 1 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian modeling of lexical knowledge acquisition in *BRAID*, a visual word recognition model

Emilie Ginestet, Sylviane Valdois and Julien Diard

1st Bayesian model of lexical knowledge acquisition

Model selects attentional displacements that maximize information gain

Model predicts serial decoding for first exposure, that transforms into parallel processing for further exposures

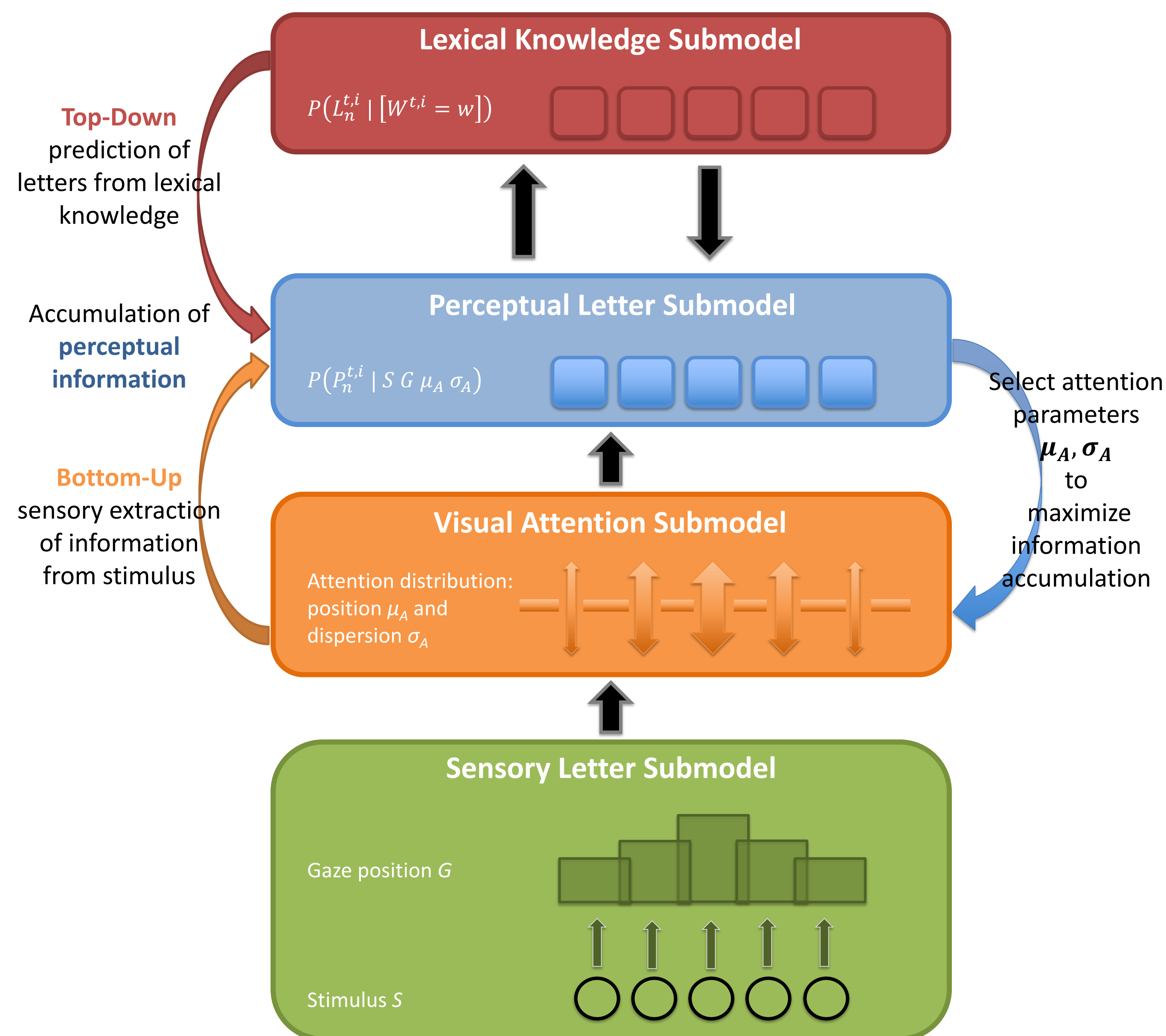
Context

Modeling lexical orthographic knowledge acquisition is one of the main current challenges. While many computational models simulate reading, word recognition, phonological transcoding or eye movement control in text reading, only one study (Ziegler, Perry, & Zorzi, 2014) has attempted to model the acquisition of lexical orthographic knowledge in children by proposing a dedicated model. This model postulates a fundamental role of phonological processing in the acquisition of lexical orthographic knowledge.

Our team has recently developed **BRAID**, a new Bayesian model of word recognition ("Bayesian word Recognition with Attention, Interference and Dynamics"), to simulate the performance of expert readers (Phenix, Valdois, & Diard, submitted; Ginestet, Phenix, Diard, & Valdois, submitted). Here, we propose an extension of BRAID by implementing a mechanism for the acquisition of new orthographic knowledge based on predictive calculations of visual-attention parameters.

Extension of the BRAID model for orthographic learning

BRAID is a hierarchical probabilistic model of visual word recognition composed of 4 submodels. We introduce a top-down dependency, so that contents of the perceptual letter submodel control attentional parameters during orthographic learning.



To model orthographic learning, we assume that visual attention displacements are chosen to optimize the accumulation of perceptual information $P(L_n^{t,i} | S G \mu_A \sigma_A)$ about letters, so as to construct efficiently the new orthographic memory trace.

Goal: accumulate perceptual information efficiently in $P(L_n^{t,i} | S G \mu_A \sigma_A)$

Acquiring lexical traces for new words

In the model, lexical knowledge is represented by probability distributions about letters, $P(L_n^{t,i} | [W^{t,i} = w])$. Their parameters are learned and updated after each exposure, by integrating the current perceptual trace $P(L_n^{t,i+1} | S G \mu_A \sigma_A)$.

$$P(L_n^{t,i+1} | [W^{t,i+1} = w]) = \frac{P(L_n^{t,i} | [W^{t,i} = w]) \times i + P(L_n^{t,i+1} | S G \mu_A \sigma_A)}{i + 1}$$

Top-Down lexical influence

Lexical knowledge about letters influences perceptual traces $P(L_n^{t,i} | S G \mu_A \sigma_A)$. This influence is modulated by the probability that the stimulus presented is a known word or not (i.e., our model of lexical decision). During orthographic learning, this influence represents a novelty detection mechanism.

Attentional displacements

We assume that visual attention parameters are controlled in order to efficiently obtain perceptual traces. This is modeled by selecting attentional parameters $(\mu_A; \sigma_A)$ that maximize a Total Gain (TG) measure. TG characterizes the information gain obtained during an attentional fixation (measured as the expected entropy gain), modulated by an estimate of the motor cost of the attentional displacement:

$$\text{Total Gain} = (1 - \alpha) \times \text{Entropy Gain} - \alpha \times \text{Motor Cost}$$

$$\Delta H(n, \mu_A^{prev}, \sigma_A^{prev}, \mu_A^{next}, \sigma_A^{next}) = H(P(L_n^{next} | S G \mu_A^{prev}, \sigma_A^{prev})) - H_{\mu_A^{next}, \sigma_A^{next}}(P(L_n^{next} | S G \mu_A^{next}, \sigma_A^{next}))$$

$$\overline{\Delta H}(\mu_A^{prev}, \sigma_A^{prev}, \mu_A^{next}, \sigma_A^{next}) = \frac{1}{n} \sum \Delta H(n, \mu_A^{prev}, \sigma_A^{prev}, \mu_A^{next}, \sigma_A^{next})$$

Finally,

$$\mu_A, \sigma_A = \underset{(\mu_A^{prev}, \sigma_A^{prev}, \mu_A^{next}, \sigma_A^{next})}{\text{max}} [(1 - \alpha) \times \overline{\Delta H}(\mu_A^{prev}, \sigma_A^{prev}, \mu_A^{next}, \sigma_A^{next}) - \alpha \times |\mu_A^{next} - \mu_A^{prev}|]$$

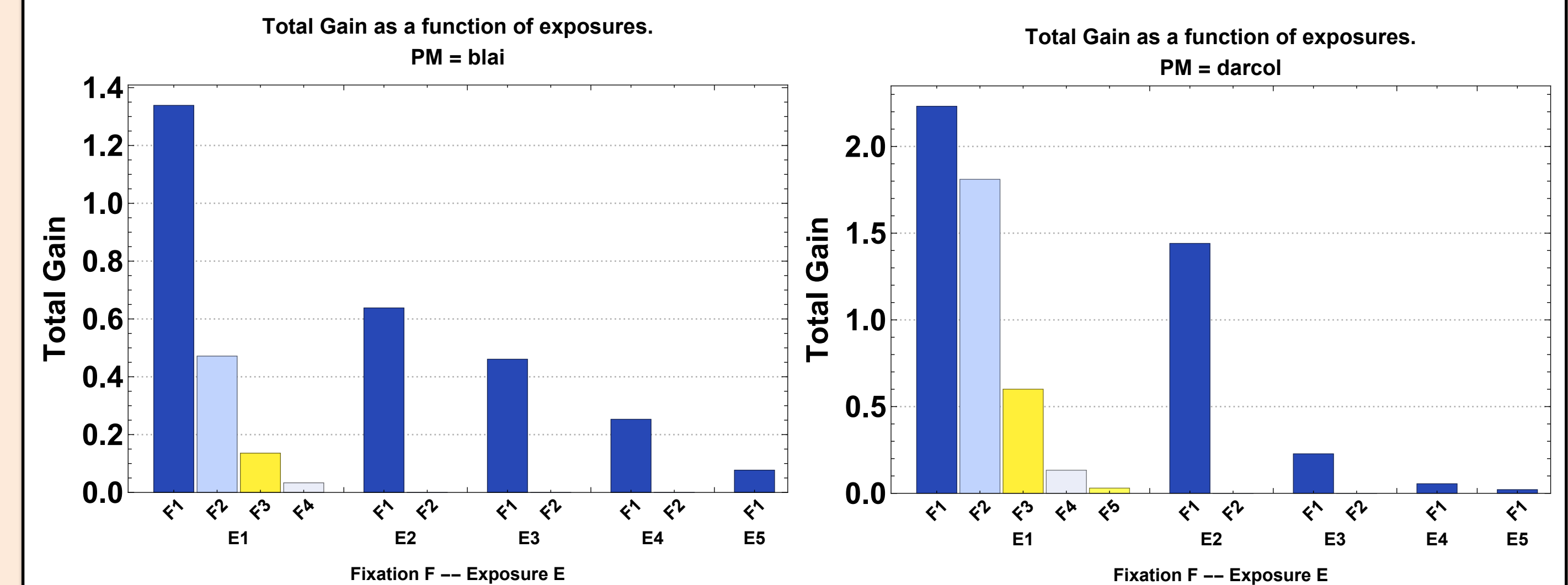
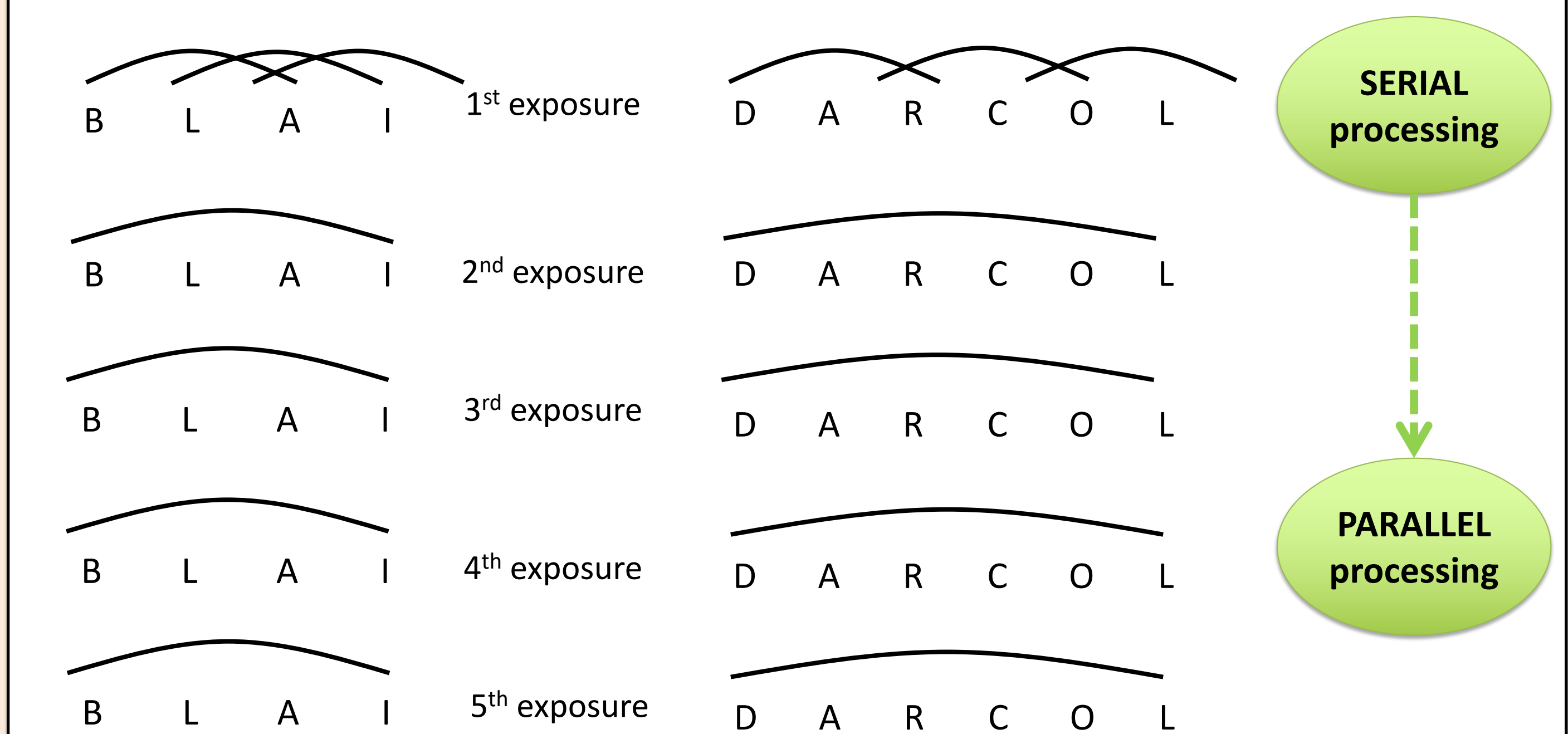
Prediction: Entropy Gain will decrease over exposures, gradually reducing the number of attentional fixations

References

Ginestet, E., Phenix, T., Diard, J., & Valdois, S. (submitted). Modelling length effect for words in lexical decision: the role of visual attention.
Phenix, T., Valdois, S., & Diard, J. (submitted). Bayesian word recognition with attention, interference and dynamics.
Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1634), 20120397.

Simulation Results

Illustration of learning two new words (4 letters, **BLAI**; 6 letters, **DARCOL**), each presented 4 times.



Summary and discussion

- Our experimental results suggest that the BRAID model successfully simulates orthographic learning of new words.
- Our model of orthographic learning is based on a strategy to optimize the accumulation of perceptual information. This strategy predicts that, for novel words, a serial decoding is the most efficient strategy, as no lexical knowledge is yet available. After few exposures, lexical knowledge complements sensory decoding so that parallel decoding becomes more efficient.
- No phonological component is involved to obtain this prediction.
- However, in our preliminary simulations, Total Gain and the number of required attentional fixations rapidly decrease. The strategy transition is fast, whatever the word length.
- We are currently conducting a behavioral experiment, with adult participants learning novel pseudo-words, in order to assess whether this transition speed is realistic.