Probabilistic modeling of orthographic learning based on visuo-attentional dynamics

Emilie Ginestet, Sylviane Valdois, Julien Diard

Psychonomic Bulletin & Review manuscript No.
(will be inserted by the editor)

Probabilistic modeling of orthographic learning based on visuo-attentional dynamics

Emilie Ginestet · Sylviane Valdois · Julien Diard

Received: date / Accepted: date

Abstract How is orthographic knowledge acquired? In line with the self-teaching hypothesis, most computational models assume that phonological recoding has a pivotal role in orthographic learning. However, these models make simplifying assumptions on the mechanisms involved in visuo-orthographic processing. Against evidence from eye movement data during orthographic learning, they assume that orthographic information on novel words is immediately available and accurately encoded after a single exposure. In this paper, we describe *BRAID-Learn*, a new computational model of orthographic learning. *BRAID-Learn* is a probabilistic and hierarchical model that incorporates the mechanisms of visual acuity, lateral interference and visual attention involved in word recognition. Orthographic learning in the model rests on three main mechanisms: first, visual attention moves over the input string to optimize the gain of information on letter identity at each fixation; second, top-down lexical influence is modulated as a function of stimulus familiarity; third, after exploration, perceived information is used to create a new orthographic representation or stabilize a better-specified representation of the input word. *BRAID-Learn* was challenged on its capacity to simulate the eye movement patterns reported in humans during incidental orthographic learning. In line with the behavioral data, the model predicts a larger decline with exposures in number of fixations and processing time for novel words than for known words. For novel words, most changes occur between the first and second exposure, that is to say, after creation in memory of a new orthographic representation. Beyond phonological recoding, our results suggest that visuo-attentional exploration is an intrinsic portion of orthographic learning, seldom taken into consideration by models or theoretical accounts.

Keywords Orthographic learning · Bayesian Modeling · Visual Attention · Eye-movements

Emilie Ginestet (✉)
Univ. Grenoble Alpes, CNRS, LPNC, 38000, Grenoble, France
E-mail: emilie.ginestet@univ-grenoble-alpes.fr

Sylviane Valdois
Univ. Grenoble Alpes, CNRS, LPNC, 38000, Grenoble, France

Julien Diard
Univ. Grenoble Alpes, CNRS, LPNC, 38000, Grenoble, France

1 Introduction

Phonological decoding – the use of spelling-sound mapping knowledge to translate letter strings into phonemes – is a first major step of reading acquisition allowing beginning readers to decode the new words they encounter while reading. The laborious and serial phonological decoding of beginning and poor readers contrasts with the fluent and immediate recognition of individual words that characterizes expert reading. Moving from slow phonological decoding to fluent reading depends on orthographic learning skills (Castles et al., 2018). However, the mechanisms by which orthographic learning occurs and how they can be modelled remain under-specified.

The self-teaching theory provided insights into one of the mechanisms at play (Share, 1995, 1999). The theory postulates that each successful decoding of a novel word provides an opportunity to learn the novel word orthographic form. Accordingly, phonological decoding is viewed as the primary cognitive mechanism involved in orthographic learning. Explicit learning of spelling-sound correspondences allows children to decode the novel word, which bootstraps orthographic knowledge acquisition. A few computational models have implemented the self-teaching mechanism (Perry et al., 2019; Pritchard et al., 2018; Ziegler et al., 2014). In these models, the phonemes corresponding to the stimulus letter-string are activated by application of grapheme-phoneme mappings, which in turn yields activation of the corresponding phonological word in long-term memory. Then, a new orthographic representation is created and the association of the new orthographic word representation with the phonological word can be learned. Ziegler et al. (2014) showed how word-specific orthographic knowledge might be successfully acquired while starting with limited knowledge of spelling-sound correspondences. Pritchard et al. (2018) showed how contextual and semantic information contributes to single word identification to facilitate irregular-word learning. However, both a force and a limit of these implementations of how children self-learn novel orthographic words is the emphasis on phonological decoding while avoiding explicit modeling of the visual mechanisms involved in novel word letter-string processing. In both computational models, a complete and immediate identification of the letters that compose the novel word is implemented, as if information on word-letter identity was fully available and memorized one-shot while reading. As acknowledged by the authors of these models themselves, such a simple one-shot approach to orthographic learning is not psychologically plausible.

First, behavioral evidence from self-teaching studies, developmental dyslexia research and animal studies suggests that word orthographic learning is not based solely on phonological decoding. Second, experimental studies using eye tracking in conditions of incidental orthographic learning clearly show that orthographic information on novel words is not immediately available but accumulates gradually in memory across successive encounters with the novel word.

Although the self-teaching theory ascribes a central role to phonological decoding in orthographic knowledge acquisition (Cunningham, 2006; de Jong et al., 2009; Kyte & Johnson, 2009; Nation et al., 2007; Share, 1999), there is evidence that orthographic learning is not fully explained by decoding ability (Castles & Nation, 2006, 2008). In particular, factors that relate to visual word processing, like "orthographic processing" and "print exposure" have been identified as contributing to the development of orthographic knowledge, beyond phonological skills (Cunningham et al., 2001; see Castles and Nation, 2006 for a review).

Further evidence against phonological processing as the unique basis of orthographic learning comes from developmental dyslexia. On the one hand, prototypical patterns of phonological dyslexia have been observed in patients who demonstrate fully developed word-specific orthographic knowledge despite major phonological deficit (Howard & Best, 1996). On the other hand, there are cases of surface dyslexia who show major deficits of irregular-word reading and spelling despite normal phonological skills (Inserm, 2007; Romani et al., 2008; Romani et al., 1999; Valdois et al., 2003). This suggests that very poor phonological decoding skills do not necessarily prevent orthographic learning and that having good phonological decoding skills does not guarantee normal development of lexical orthographic knowledge. Of particular interest for the present purpose, search for the cognitive deficits associated with developmental surface dyslexia revealed that a selective orthographic deficit was associated with a deficit of the simultaneous processing of distinct visual elements, dubbed the visual attention (VA) span deficit (Bosse, 2005; Dubois et al., 2010; Valdois et al., 2003). Further evidence that VA span more specifically relates to reading subskills that reflect word-specific orthographic knowledge – like irregular word reading (Bosse & Valdois, 2009), reading speed (Lobier et al., 2013; van den Boer et al., 2015; van den Boer & de Jong, 2018) or the length effect in word reading (van den Boer et al., 2013) – supports a potential contribution of VA span to word-specific orthographic knowledge acquisition. More direct evidence comes from studies showing a link between VA span and spelling acquisition (Niolaki et al., 2020; van den Boer et al., 2015) and from studies showing that VA span modulates novel word orthographic learning (Bosse et al., 2015; Chaves et al., 2012; Ginestet et al., 2020; Marinelli et al., 2020). Without minimizing the role of phonological skills in orthographic acquisition, these findings suggest that visual factors independently contribute to the development of word-specific orthographic knowledge. Data from animal studies further suggest that the contribution of visual processing skills to orthographic knowledge acquisition may have been underestimated since animals can acquire impressive orthographic knowledge in the absence of language and phonological skills (Grainger et al., 2012; Rajalingham et al., 2020; Scarf et al., 2016). These findings highlight the urgency to better understand how visual processing and visual attention skills contribute to orthographic learning and self-teaching.

Finally, recent exploration of eye movements in conditions of novel word incidental learning revealed that orthographic learning is modulated by complex visual processes. While incidental learning begins from the first encounter with the novel word (Bosse et al., 2015; Bowey & Muller, 2005; Cunningham, 2006; Nation & Castles, 2017; Share, 1999, 2004; Tucker et al., 2016), orthographic learning is not completed at the end of the first exposure but requires multiple encounters. Strong variations in eye movements due to repeated exposure with the same novel word are reported across the first two or three exposures, but learning effects can be observed later on and five successive exposures can be insufficient for the novel word (or rare words) to be processed as a known word (Ginestet et al., 2020; Joseph & Nation, 2018; Joseph et al., 2014; Nation & Castles, 2017; Pellicer-Sanchez, 2016). Clearly, orthographic processing affects incidental learning over multiple exposures; this contrasts with the simplified picture assumed by most computational models. Monitoring eye movements provided additional insights on the mechanisms at play.

Gradual decrease in processing time (gaze duration and fixation duration) with successive encounters is the main indicator of orthographic learning. The reduction in processing time across exposures, as assessed by measuring eye movements, is associated with increased performance on offline measures of novel word spelling knowledge. This suggests that letter identification is boosted from exposure to exposure through top-down influence due to gradual reinforcement of the novel word orthographic representation (see Ginestet et al., 2020; Joseph and Nation, 2018; Joseph et al., 2014 and, for qualitatively consistent observations, see, Pagan and Nation, 2019). Available data thus suggests that orthographic learning is a gradual, not an all-or-nothing, process that relies on close interactions between bottom-up processing for the extraction of letter information from the novel printed word and top-down lexical influences, including the influence of the orthographic representation of the novel word currently being acquired.

Overall, current models of the self-teaching mechanism implement orthographic learning as a one-shot process allowing the immediate and accurate memorization of the whole orthographic form of a novel word as far as it has been accurately decoded and phonologically recognized. In contrast, behavioral data from eye movement studies show that the oculomotor pattern evolves across repeated exposures to the same novel word suggesting a gradual, not one-shot, acquisition of orthographic knowledge. Furthermore, additional behavioral data suggest that, beyond phonology, visual attention might be involved in orthographic learning. Unfortunately, no current computational model implements all the mechanisms required to predict the evolution of eye movements during orthographic learning. On the one hand, models of reading acquisition do not incorporate any of the mechanisms of visuo-orthographic processing that are postulated by models of orthographic word recognition. In particular, reading acquisition models do not implement the mechanisms of inter-letter visual similarity and lateral interference that are critical in word recognition models. On the other hand, models of eye movement control implement the visual acuity and visual attention components required to account for eye movements in reading but they incorporate none of the visuo-orthographic processes that are central for word recognition models and no mechanism of orthographic learning.

Our main contribution in the present study was to implement a more integrated computational model and assess its ability to predict the evolution of eye movements during orthographic learning. For this purpose, we started from a recently developed word recognition model, the *BRAID* model (for *Bayesian model of word Recognition with Attention, Interference and Dynamics*; Phénix, 2018; Phénix et al., submitted), which includes not only the mechanisms of visual letter similarity and lateral interference classically found in word recognition models, but further the mechanisms of visual acuity and visual attention that are typical of eye movement control models. We extended the *BRAID* model by adding learning mechanisms. As a result, the extended model, called *BRAID-Learn*, features simultaneously the properties of an efficient word recognition model, some of the processes involved in eye movement control, and the mechanisms required for orthographic learning. We then used the *BRAID-Learn* model to predict the evolution of eye movement patterns when being repeatedly exposed to the same set of novel words. A main challenge here was to use the same set of default parameter values that was previously used to simulate a variety of word recognition effects (like frequency and neighborhood effects (Phénix, 2018; Phénix et al., submitted; Phénix et al., 2018), the word superiority or the OVP effect (Phénix, 2018; Phénix et al., submitted; Valdois et al., submitted), or word length effects (Ginestet et al., 2019; Saghiran et al., 2020)), in an attempt to account for word recognition, orthographic learning and eye movement data in a single computational framework. Therefore, overall, our main objective was to explore to what extent a model that was not specifically designed to account for eye movement control while reading would generalize and predict the evolution of eye movement patterns during the orthographic learning of novel words.

The rest of this paper is structured as follows. First, we propose a brief description of the *BRAID* word recognition model and describe the three mechanisms of orthographic learning that were implemented to develop the *BRAID-Learn* model. Second, we focus on an example to provide an in-depth illustration on how the learning mechanisms affect the processing of known words and novel words. Last, we confront the *BRAID-Learn* model to a set of known and novel words to evaluate its capacity to predict the eye movement patterns that characterize the orthographic acquisition of new words by humans while reading.

2 The *BRAID-Learn* model

In this section, we describe the *BRAID-Learn* model, as an extension of the *BRAID* word recognition model. Since both models are nested, we first provide a brief description of the *BRAID* model and, second, we present the mechanisms added to *BRAID* to model orthographic learning.

2.1 The *BRAID* model

A full description of the *BRAID* model is provided elsewhere (Phénix, 2018; Phénix et al., submitted), and beyond the scope of this paper. Instead, we briefly describe some salient features of the *BRAID* model that are relevant to understanding the proposed extension, *BRAID-Learn*.

In a nutshell, *BRAID* is a probabilistic, hierarchical model of visual, attentional and lexical knowledge that allows simulating tasks such as letter recognition, word recognition and lexical decision. The *BRAID* model can be seen as building upon the three-layer architecture of previous models and extending them. In particular, the *BRAID* model features an original visual attention layer, that modulates letter and word perception.

Mathematically, *BRAID* is defined by a joint probability distribution, linking sensory, perceptual and lexical probabilistic variables. This joint probability distribution is defined thanks to conditional independence hypotheses, which allow delineating five submodels and their connections; this forms the structure of the model (see Fig. 1). We now describe some features of each five submodels of the *BRAID* model, and how *BRAID* is then used, thanks to Bayesian inference, to simulate letter recognition, word recognition and lexical decision.

2.1.1 The four submodels of the BRAID model

The "Letter Sensory" submodel This submodel concerns low-level visual processing of letter stimuli, S_1^1 to S_N^T, with subscripts 1 to N referring to spatial positions, and superscripts 1 to T referring to time instants (we will use $S_{1:N}^{1:T}$ as a shorthand for the whole set of these variables). From the stimulus, this

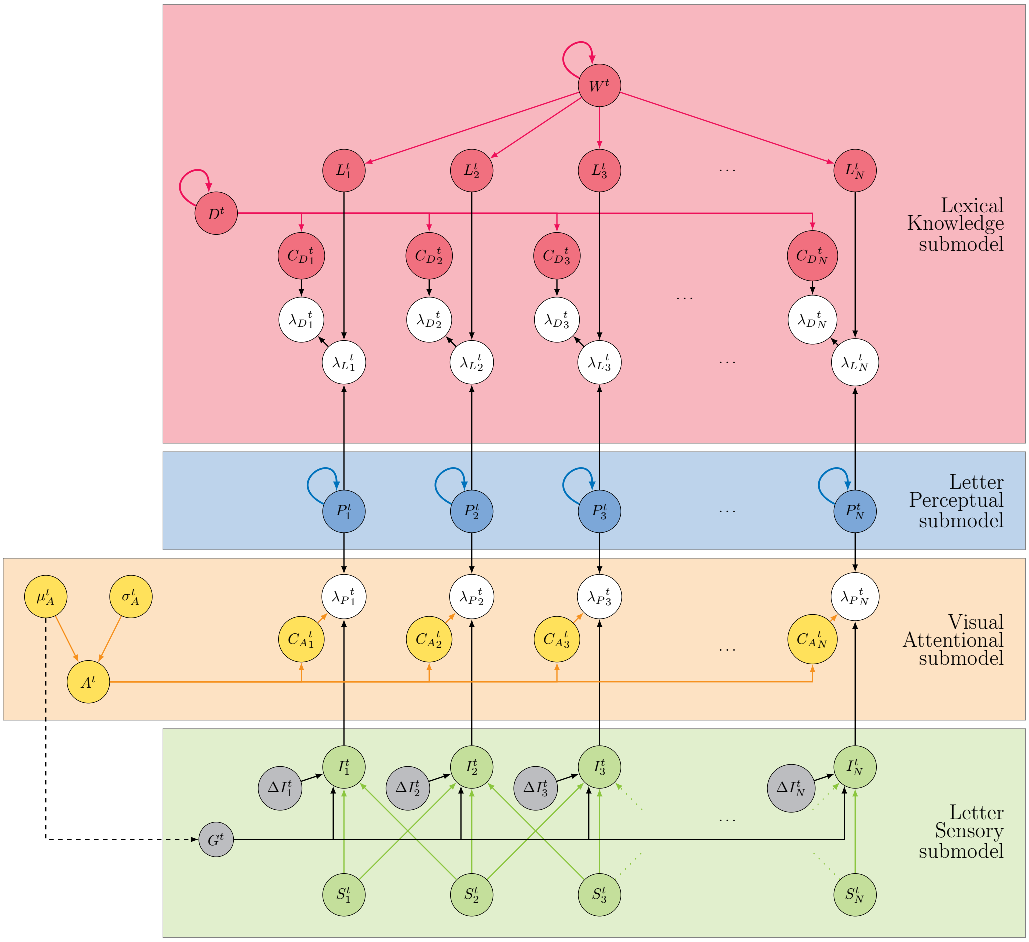

Fig. 1 Graphical representation of the structure of the *BRAID* model. Each of the four colored blocks represents a submodel; each node of the graph represents a variable of the model; and each arrow represents a probability distribution of the model. The graphical schema presented here corresponds to a time-slice, at time instant T (note the superscripts $T-1$ and T in some nodes) of the *BRAID* model configured for a 5-letter stimulus (note the subscripts from 1 to 5, in variables such as S_1^T to S_5^T). See text for details.

submodel essentially infers "internal" representations of letter identity, in the form of discrete probability distributions, over variables $I_{1:N}^{1:T}$, with their domain the set of the 27 possible characters (26 letters plus a special character denoting an unknown or missing letter).

The letter sensory submodel includes a confusion matrix, from stimuli to internal representations of letters, calibrated to match typical, expert reader performance in isolated letter recognition (Geyer, 1977). Several mechanisms modulate letter recognition at the sensory level. Gaze position within the input letter string is implemented (with variable $G^{1:T}$) together with an acuity gradient that increases uncertainty on letter identification as a function of eccentricity from gaze position. A mechanism of lateral interference from adjacent letters contributes to uncertainty on letter identity and letter position, yielding crowding effects.

The "Visual Attentional" submodel Using intermediate variables and probability distributions – technically, so called "coherence" (Bessière et al., 2008; Gilet et al., 2011) and "control" variables (Phénix, 2018) –, the visual attentional submodel acts as a layer filtering the transfer of bottom-up information, i.e., from the "Letter Sensory" submodel to the "Letter Perceptual" submodel. This allows to modulate

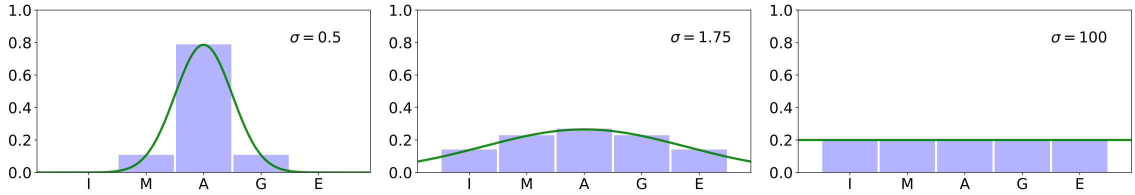

Fig. 2 Illustration of the attention distribution over the letter string of the stimulus word IMAGE for a position of attention $\mu_A = 3$, and for different values of attention dispersion σ_A. The y-axis represents the attention quantity for each position, and the x-axis represents letter positions. Left: $\sigma_A = 0.5$; Middle: $\sigma_A = 1.75$; Right: $\sigma_A = 100.0$.

letter information transfer differently for each position, depending on visuo-attentional distribution. To do so, the probability distribution $P(A^t \mid \mu_A^t\ \sigma_A^t)$ at time t characterizes the spatial distribution of visual attention by a discretized and truncated Gaussian probability distribution. Its mean μ_A^t represents the position of the attentional focus (which we assume, in all the simulations presented here, to coincide with gaze position G^t), and its standard deviation σ_A^t represents attentional dispersion.

As Fig. 2 shows, the smaller the value of σ_A, the more attention is focused on a small number of letters. For instance, with $\sigma_A = 0.5$, attention is focused, enhancing the perceptual accumulation of information about the 3rd letter, mostly (in our example, $\mu_A = 3$ and the stimulus is 5-letter long), to the detriment of external letters (e.g., the 1st and 5th are hardly processed). On the other hand, a large value of σ_A (for example, 100) simulates a uniform distribution of attention over the stimulus. In this case, the speed of perceptual information accumulation is equal for all letter positions. Finally, with $\sigma_A = 1.75$, the attention distribution allows to slightly modulate the information transfer speed over the five letters, in this example favoring the processing of central letters. The 1.75 value for attention dispersion σ_A is the default value, calibrated from independent data (Ginestet et al., 2019) from lexical decision mega-study (Ferrand et al., 2010).

The "Letter Perceptual" submodel The third submodel we describe is the letter perceptual submodel, in which evidence about letter identity is accumulated, over time, into probabilistic variables $P_{1:N}^{1:T}$. It can be seen as a series of Markov chains, one for each position n. Each such Markov chain, in essence, is a temporally evolving probability distribution, here over the discrete space of all 27 possible characters. This probabilistic model both has intrinsic dynamics, according to which information gradually decays towards a resting state which is the uniform distribution, and input information from "neighboring" submodels (i.e., those linked to it by probabilistic dependencies, see Fig. 1). In the *BRAID* model, the letter perceptual submodel receives, on the one hand, perceptual information from the letter sensory submodel filtered by the visual attentional submodel, in a bottom-up manner, and on the second hand, lexically predicted information from the lexical knowledge submodel, in a top-down manner.

The "Lexical Knowledge" submodel This submodel encodes, into the model, knowledge about a set of known words W, i.e., a lexicon. Over this space, a temporal model, again akin to a Markov chain, is defined. The initial state of this temporal model is the prior probability distribution $P(W^0)$, that encodes the frequency of words of W, as in the Bayesian Reader model (Norris, 2006). The intrinsic dynamics of the distribution over W, as above for P, is also a gradual decay towards the initial state.

Words (w in W) are associated with their corresponding letter sequence $L_{1:N}^{1:T}$ by a probabilistic model, such that, in each position, the correct letter at that position for this word has a high probability value (0.974), and all other alternatives have small probability values (0.001).

Finally, a third and final Markov chain, over variable D, might be interpreted as a "lexical membership and word familiarity check". Variable D is Boolean, with the "True" value representing that a word stimulus belongs to the known lexicon. The initial, prior distribution $P(D^0)$ is uniform, representing a 50/50 chance that the input stimulus is a known word (a viable assumption to simulate many experimental setups, although surely not realistic in ecological situations). Variables $D^{1:T}$ are related to Boolean variables $C_{D}{}_{1:N}^{1:T}$, in a probabilistic model that represents knowledge about whether a sequence of stimulus letters corresponds to a known word, or not: for a known word, all variables $C_{D}{}_{1:N}^{1:T}$ are assumed to be "True"; on the contrary, for a sequence of stimulus letters that is not a known word, at least one of the variables $C_{D}{}_{1:N}^{1:T}$ is assumed to be "False". These patterns of values serve as templates, to be compared with values of the coherence variables between the perceptual evidence about letter identity $P_{1:N}^{1:T}$ and letter sequence $L_{1:N}^{1:T}$, so that "observing" the flow of information between these two variables allows to infer whether the input stimulus is a known word or not.

2.1.2 Probabilistic questions to simulate cognitive tasks

The *BRAID* model expresses, using probability distributions, knowledge related to letter identity, how known words are related to their corresponding letter sequences, and how to describe whether a sequence of stimulus letters corresponds to a word of the known lexicon. This knowledge is then used in several cognitive processes, which we simulate by computing probabilistic distributions of interest using Bayesian inference. We call this "asking a probabilistic question" to the model.

For instance, the first cognitive task we consider is letter recognition. It is modeled by the following probabilistic question:

$$Q_{P_n^T} = P(P_n^T \mid [S_{1:N}^{1:T} = s]\ [G^{1:T} = g]\ \mu_A^{1:T}\ \sigma_A^{1:T}\ [\lambda_P{}_{1:N}^{1:T} = 1]\ [\lambda_L{}_{1:N}^{1:T} = 1]) \ , \tag{1}$$

which can be read as: What is the probability distribution over the perceived letter at position n, at time step T, given the stimulus letter sequence s, gaze position g, the current attentional distribution (μ_A, σ_A), and given that information is allowed to propagate from the stimulus to the lexical submodel ($[\lambda_P{}_{1:N}^{1:T} = 1]$, $[\lambda_L{}_{1:N}^{1:T} = 1]$)?

For lack of space, we do not provide here the mathematical expression that Bayesian inference yields as an answer to this question (Phénix, 2018). However, the resulting computation can be interpreted as in classical, three-layer models with lexical, top-down influence: the sensory letter submodel extracts information about letter identity from the sequence stimulus; part of this perceptual information, depending on the attentional distribution, is propagated and accumulated into the dynamic models of the perceptual layer submodel. These propagate to the lexical submodel, gradually changing the probability distribution over words which, in a feedback manner, informs the perceptual layer submodel.

Fig. 3 (top) illustrates the temporal accumulation of perceptual information about letters composing the 8-letter long French word *MENSONGE* (*LIE*), in a simulation where the eye position g and visual

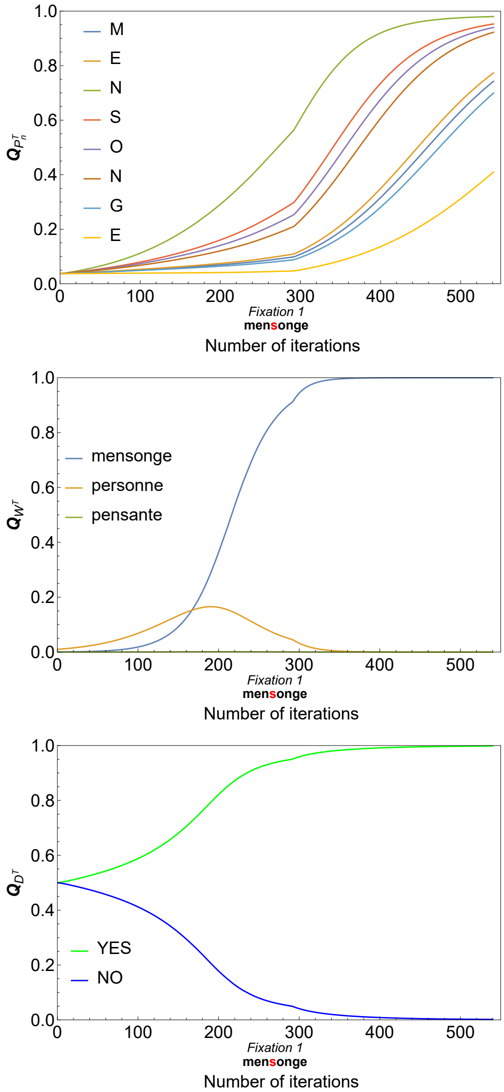

Fig. 3 Evolution of inference for $Q_{P_n^T}$ (top; Eq. (1)), Q_{W^T} (middle; Eq. (2)) and Q_{D^T} (bottom; Eq. (3)) as a function of simulated time (x-axis) for the 8-letter stimulus *MENSONGE* (*LIE*), with $g = \mu_A = 4$ (eye and attention are positioned over letter "S", indicated in red under each plot) and $\sigma_A = 1.75$ (default value for attentional dispersion). For letter recognition (top plot), only the probability value of the correct letter at each position is shown. For word recognition (middle plot), only the probability values of the three most probable words are shown (note that the third more probable competitor, word *PENSANTE* (*THINKING*), is very close to 0, almost superposed with the x-axis). For lexical decision, for each time-step, the whole probability distribution over the Boolean lexical membership variable D^T is shown.

attention focus position μ_A are assumed to be on the fourth letter ("S") for the whole simulation (and with attention dispersion at its default value, $\sigma_A = 1.75$). We see that perceptual information gradually accumulates towards the correct recognition of all letters, but that it does so slower as distance to the position of the eye and of the attentional focus increases (i.e., in this example, faster for central letters "N", "S" and "O" under the attention focus than for external letters, that is, the initial "M" and the final "E").

The second task, word recognition, is modeled in a similar manner, by considering the probabilistic question:

$$Q_{W^T} = P(W^T \mid [S_{1:N}^{1:T} = s]\ [G^{1:T} = g]\ \mu_A^{1:T}\ \sigma_A^{1:T}\ [\lambda_{P}{}_{1:N}^{1:T} = 1]\ [\lambda_{L}{}_{1:N}^{1:T} = 1])\ . \tag{2}$$

Contrary to letter recognition, in word recognition the "target space", that is to say, the domain of the probability distribution of interest, is the word space W. The result of inference, in this case, is similar to the inference for letter perception, with the same flow of information, from the stimulus, up to the lexical submodel, with a feedback to the letter perception submodel.

Coming back to the example of processing the stimulus *MENSONGE*, simulation of word recognition leads to the progressive activation of the corresponding word of the lexical space ($W = \mathit{MENSONGE}$) and its lexical competitors, such as $W = \mathit{PERSONNE}$ (PERSON) and $W = \mathit{PENSANTE}$ ($\mathit{THINKING}$), as shown in Fig. 3 (middle). Comparing letter recognition and word recognition (respectively, top and middle plots of Fig. 3) shows that the probability converges in word space faster than in letter space; in other words, assuming identical decision thresholds for words and letters would yield faster word recognition than letter recognition: the word would be recognized faster than its letters. Such an observation is consistent with human observations (Phénix, 2018).

The third and final cognitive task is lexical decision, that is to say, recognizing whether the input letter sequence matches that of a known word. The probabilistic question is:

$$Q_{D^T} = P(D^T \mid [S_{1:N}^{1:T} = s] \; [G^{1:T} = g] \; \mu_A^{1:T} \; \sigma_A^{1:T} \; [\lambda_{D}{}_{1:N}^{1:T} = 1] \; [\lambda_{L}{}_{1:N}^{1:T} = 1]) \; . \tag{3}$$

As previously, a stimulus is given, gaze position and attention distributions are set, and information is allowed to propagate into the model. However, here, we do not assume that there is a match between the stimulus and a known word; instead, by involving the lexical membership variables ($\lambda_{D}{}_{1:N}^{1:T} = 1$), the probability distribution over variables $\lambda_{L}{}_{1:N}^{1:T}$ is evaluated, in essence, performing error detection in the stimulus with respect to all possible known words. Here, information flows through the whole *BRAID* architecture: as previously, from the stimulus to the lexical submodel and back down to the perceptual letter submodel, with the added involvement of the lexical membership variable D^T as an observer.

We reprise once more our example where the stimulus *MENSONGE* is processed. Fig. 3 (bottom) illustrates the evolution of the probability distribution over variable D^T as a function of time: we observe that the probability that D^T is YES increases steadily, so that the model correctly identifies the input stimulus ($W = \mathit{MENSONGE}$) as a known word.

2.2 The *BRAID-Learn* model

The *BRAID-Learn* model is an extension of the *BRAID* model, that incorporates three new mechanisms allowing learning the orthographic representations of visually presented new words. Its main assumption is that the model's aim is to accumulate efficient information about letters of the stimulus, so that, when faced with a novel word, this information can be learned as an orthographic trace paired with a newly allocated point of the lexicon W. Therefore, the three main mechanisms of the *BRAID-Learn* model concern how it accumulates information about letters, how novelty detection influences stimulus processing, and, finally, how the resulting perceived traces are used to learn a new orthographic trace or reinforce an already existing one. Fig. 4 shows a graphical representation of the *BRAID-Learn* model (to compare with the *BRAID* model, see Fig. 1).

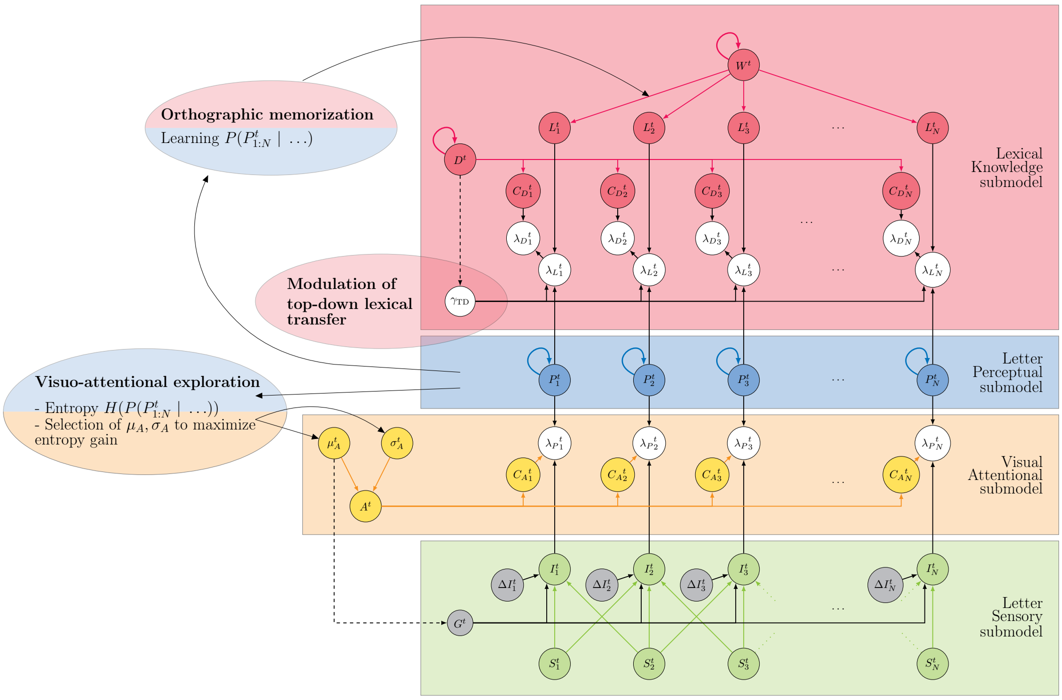

Fig. 4 Graphical representation of the *BRAID-Learn* model. The four colored blocks are the same submodels as in the *BRAID* model, with the same graphical convention (see Fig. 1). To this architecture, the *BRAID-Learn* model adds three mechanisms, represented as colored ovals, that transfer and transform (colored and black arrows going through the ovals) information contained in portions of the *BRAID* model.

2.2.1 Efficient accumulation of perceptual evidence about letters

To model the accumulation of perceptual evidence about letters in a stimulus sequence, we consider the letter recognition task of Eq. (1). It is defined as a function of the current visual and visuo-attentional parameters, namely gaze position g^T, the position of the attentional focus μ_A^T and the dispersion $\sigma_A^{1:T}$ of the visuo-attentional distribution. Of course, fixing a unique attentional distribution and gaze position throughout stimulus processing can yield inefficient processing. For long words (e.g., 8-letter long), concentrating attention leaves almost no perceptual processing available for some letters, and spreading attention maximally (i.e., distributing attention uniformly) yields massive, unrealistic length effects (Ginestet et al., 2019). Furthermore, and as described previously, it is well-known that eye-movements are observed in natural settings, for instance for long words and during new word processing (Lowell & Morris, 2014).

Therefore, the first and main mechanism of *BRAID-Learn* is a visuo-attentional control mechanism, that is to say, the model controls and changes its attentional distribution and gaze position over time, so as to accumulate perceptual evidence efficiently. To describe the sequencing of several fixations, we refine our temporal notation. A simulation from time-steps 0 to T is broken down as a series of exposures to a stimulus letter sequence, e from 1 to E, each exposure consisting of a variable number of fixations f from 1 to F and each fixation being of variable length, from 1 to T_f time-steps (see Fig. 5).

During one exposure, at the end of each fixation, the model selects the attentional distribution parameters that would provide the most efficient accumulation of perceptual evidence to be yet gathered. The

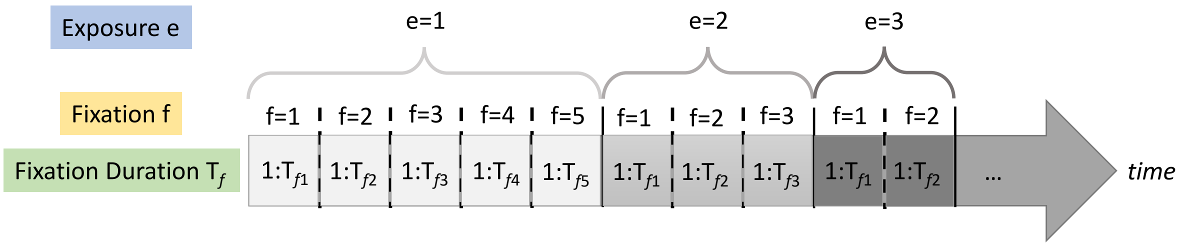

Fig. 5 Schematic representation of the time course of a sequence of several fixations and exposures. Each time a word is encountered (each exposure e), it is fixated once or more times (fixations f), and each such fixation consists in a (variable) duration T_f during which the eye position and the visuo-attentional distribution parameters are fixed. This results in a varying total processing time for each exposure (sum of all T_fs).

classical mathematical measure of the information content of a discrete probability distribution $P(X)$ is its entropy, noted $H(P(X))$ and defined by:

$$H(P(X)) = -\sum_X \left(P(X) \log P(X)\right) \ . \tag{4}$$

The lower the entropy of a probability distribution, the more it contains information: for a given variable X, entropy is maximal for the uniform distribution over X, which encodes maximal uncertainty, and 0 for Dirac distributions, which encode maximal certainty. Therefore, decreasing entropy amounts to gaining information.

The *BRAID-Learn* model aims at optimizing information gain by maximizing entropy decrease. Mathematically, before fixation $f+1$, we enumerate a range of possible values for upcoming attention position $\mu_A^{T,f+1,e}$ and dispersion $\sigma_A^{T,f+1,e}$; for each such possible future attention distribution, and assuming that the input stimulus will not change during next fixation, we simulate letter recognition in each position n with

$$\begin{aligned} &P_{\text{next}}(n, \mu_A^{f+1,e}, \sigma_A^{f+1,e}) \\ &= P(P_n^{T,f+1,e} \mid [S_n^{T,f+1,e} = s]\ [G^{T,f+1,e} = g^{f+1,e}]\ \mu_A^{T,f+1,e}\ \sigma_A^{T,f+1,e}) \ . \end{aligned} \tag{5}$$

Recall that we assume that gaze position and attention position coincide, so that $g^{T,f+1,e} = \mu_A^{T,f+1,e}$. We can then compute the entropy gain between the predicted and current distribution over letters, for all possible attention distribution parameters and average it across positions; we note this $\overline{\Delta H}(\mu_A^{T,f+1,e}, \sigma_A^{T,f+1,e})$.

To model the physical "motor cost" of performing the visuo-attentional displacement to each enumerated future fixations, we use a straightforward measure, considering only the magnitude of the supposed displacements of gaze and attention: $MC(\mu_A^{T,f+1,e}) = |\mu_A^{T,f+1,e} - \mu_A^{T,f,e}|$. We use this measure to penalize large displacements of gaze and attentional positions, so that the overall gain measure TG that the model maximizes is a weighted combination of information gain penalized by motor cost:

$$\overline{TG}(\mu_A^{T,f+1,e}, \sigma_A^{T,f+1,e}) = (1-\alpha)\overline{\Delta H}(\mu_A^{T,f+1,e}, \sigma_A^{T,f+1,e}) - \alpha MC(\mu_A^{T,f+1,e}) \ . \tag{6}$$

Finally, the model selects, for its next fixation, attentional parameters and gaze position that maximize measure $\overline{TG}$.

Having described how, at any point in time, the next fixation parameters are selected, we define the initial parameters and termination criterion. Whatever the stimulus, whether it is a word or not, and since the model, at initialization, has no knowledge of the stimulus type, the parameters for the first fixation are identical.

Therefore, in the context of current experiments, that only deal with 8-letter long stimuli, we assume that gaze and attention “land” at position $\mu_A^{T,1,e} = 3$ whatever the exposure e. This initial position is the rounded value closest to the one from our previous experimental observations (3.01 across all item types, i.e., for words and pseudowords, and across all repetition exposures) in which expert readers had to read 8-letter words and pseudowords (Ginestet et al., 2020). For the initial dispersion of visual attention, we apply the usual default value in the *BRAID* model: $\sigma_A^{T,1,e} = 1.75$.

We define two termination criteria: the first defines how long each fixation is going to last, and the second is used to decide that no further fixations are going to be performed. Concerning fixation duration, we assume that the model aims at having as short fixations as possible (later on, during data analyses, back-to-back fixations on the same spatial position are aggregated and counted as a single fixation on this position; aiming for short fixations is not a theoretical claim, instead it just yields temporal granularity in our simulations). Initial simulations have shown that, in the first few iterations of the predictive evaluation of entropy gain, the “winning parameters” were numerically close, until a clear set of value emerged and, most of the time, stayed ahead until the maximal window of predictive computation. This maximal time is set, for current experiments, at $T = 290$ iterations, well above the average fixation duration reported for novel words in behavioral experiments (Ginestet et al., 2020; Joseph et al., 2014; Pellicer-Sanchez, 2016). We thus detect the time-step T_f for which the predicted winning parameter values have been stable for 20 previous time-steps. Finally, we set the minimal duration T_f to be at 50 iterations. The upcoming fixation is then performed with these winning parameters for that duration.

The second termination criterion prevents a further fixation when its expected information gain is below a threshold. Since fixations are of varying duration T_f, this is scaled as a function of T_f. We have empirically calibrated our stop criterion to correspond to 1 nat of information gain for the whole word, that is, $1/N$ nats for $\overline{\Delta H}$ (recall that N is the length of the input word), for a fixation of 250 iterations (our simulations use natural base e for entropy calculation, which is therefore measured in nats instead of bits). Therefore, our termination threshold is $T_f/(N \times 250)$; whenever a fixation is selected and associated to an information gain below this threshold, it is not performed by the model, and the current exposure e is considered terminated.

2.2.2 Modulation of lexical influence during word learning

The second main ingredient of the *BRAID-Learn* model is a mechanism to modulate the amount of top-down lexical information during word processing, as a function of word familiarity. An algorithmic description of the desired mechanism is as follows: if the input letter sequence is a known word, then strong top-down lexical information can be fed back to the letter perceptual sub-model, to speed up letter identification, in turn speeding up word recognition. On the other hand, if the input letter sequence is not a known word, then top-down lexical information should be diminished to try to avoid generalization toward the closest word in the lexicon, as it would yield illusory letter percepts, resulting in failure to veridically process the letters of the input novel word.

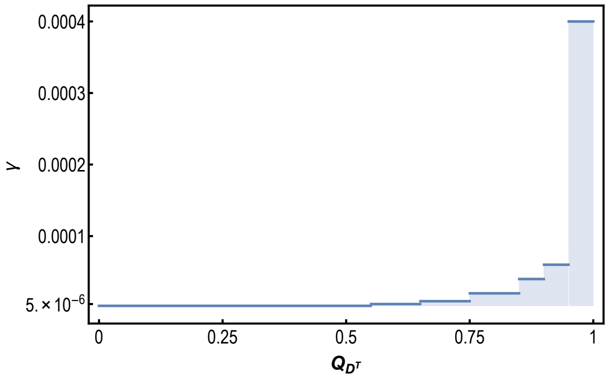

Fig. 6 Graph of the function of parameter γ (y-axis), that pilots the amount of top-down transfer of lexical information, as a function of Q_{D^T}, the probability that $D^T = \text{true}$ (x-axis), computed by Eq. (3), that represents probability of lexical membership and word familiarity.

For the sake of brevity, we do not describe here the probabilistic model that allows modulating the top-down influence from the lexical knowledge sub-model to the letter perceptual model in *BRAID-Learn*. It involves building an asymmetric layer of coherence variables between these sub-models, and piloting, via control variables, the amount of information propagating top-down; this mechanism is mathematically similar to how we control, in the visual attentional sub-model, the amount of information propagating bottom-up from the letter sensory submodel to the letter perceptual sub-model. We note γ the parameter introduced by this mechanism; the higher γ, the more there is top-down lexical information transfer.

Finally, we modulate γ as a function of how likely it is that the input letter sequence corresponds to a known word. In the model, this information is already represented, by the probability distribution over the lexical membership variable D^T. Piloting γ as a function of D^T can be interpreted as using the "lexical decision" variable space to modulate lexical influences over letter perception. Note that this does not mean that lexical decision is performed per se, as no decision threshold is involved, and the task does not consist in deciding whether the input is a word or not; instead, we assume that lexical membership is assessed in an on-going manner, even during letter and word recognition, and modulates the information flow of these tasks, at each instant. Here, the probability distribution over D^T can be interpreted as an online evaluation of lexical membership and of word familiarity.

To define the mathematical relationship between D^T and γ, our main theoretical assumption is that top-down lexical influence increases for familiar words. In mathematical terms, this results in γ being a monotonously increasing function of the probability that $D^T = \text{true}$. Furthermore, empirical exploration shows that γ needs to have small values; the lexical knowledge model contains a lot of information (it is of low entropy, as it consists of almost-Dirac distributions) and injecting it too fast into the letter perceptual letters results in trumping sensory evidence by lexical feedback. For instance, when $\gamma = 1$, and whatever the input letter sequence, the probability distributions over letters at the perceptual layer converge towards the letters of the most frequent word of the lexicon in a few iterations. We chose to implement the relation giving γ as a function of the probability that $D^T = \text{true}$ (as evaluated by Eq. (3)) by a piece-wise, monotonously increasing constant function, shown Fig. 6.

We note that the chosen function includes a sudden increase for γ when the probability that $D^T = \text{true}$ passes .95. When γ increases in such a manner, this increases the top-down lexical influence, so that the probability distribution over letters P_n^t suddenly receives more lexical evidence. In our simulations, this results in noticeable increases in the slopes of curves representing the evolution of probabilities for letters, words and lexical membership (e.g., see Fig. 3 at iteration $t = 291$).

2.2.3 Memorization and update of orthographic traces

Finally, a third mechanism allows updating lexical knowledge; this is the last step in the learning process of *BRAID-Learn*. It takes effect once an exposure is considered terminated, that is, once one of the termination criteria of visuo-attentional exploration is satisfied. The lexical knowledge sub-model is updated to learn the perceived letters, either integrating them into the already available probabilistic model for that word, if it was already known, or using them to create a new lexical trace, if the input sequence was detected as a new word by the lexical decision process (Eq. (3)).

In the first case, that is, for updating a lexical distribution, at the end of exposure e, and for each position n, the complete probability distribution about the perceived letter, $P(P_n^{T,f,e} \mid [S_n^{T,f,e} = s]\ [G^{T,f,e} = g]\ \mu_A\ \sigma_A)$, is combined with the previous probability distribution about the letter at that position, $P(L_n^e \mid [W^e = w])$, in the lexical sub-model, for the recognized word w:

$$\begin{aligned} &P(L_n^{e+1} \mid [W^{e+1} = w]) = \\ &\Big[P(L_n^e \mid [W^e = w]).(e-1) + P(P_n^{T,f,e} \mid [S_n^{T,f,e} = s]\ [G^{T,f,e} = g]\ \mu_A\ \sigma_A)\Big]/e \end{aligned} \tag{7}$$

The model also increments by 1 (arbitrarily) the estimated frequency count of word w, in the prior probability distribution of the lexical sub-model.

In the second case, that is, for creating a new lexical distribution when the input letter sequence was recognized as a new word by lexical decision, a new entry w_{new} is allocated in word space W, and the initial letter trace for that word is simply the probability distributions over its perceived letters after this first exposure.

2.3 Summary

The *BRAID-Learn* model includes three mechanisms that affect letter identification within strings during word recognition, namely an acuity gradient, a mechanism of lateral interference between adjacent letters and a visual attention filter. The model assumes that a novel word trace is created each time the input letter-string is detected as not belonging to the model lexical knowledge. Furthermore, detecting that the input is novel entails decreasing the top-down feedback from word knowledge to letter perception; this yields a relative increase in the effect of perceptual evidence about letters from bottom-up processing. In other words, bottom-up information is privileged as the principal source of information on letter identity. Visuo-attentional exploration during processing is defined by a mathematical principle of entropy gain maximisation. The entropy gain maximisation principle allows selecting the visuo-attentional distribution parameters – attentional focus and dispersion – more likely to speed-up accumulation of perceptual information about letters. This mechanism leads the model to realize as many visuo-attentional displacements as necessary as long as perceptual information is not precise enough. Visuo-attentional exploration is further constrained by a motor-cost parameter that penalizes large displacements over the letter-string. When visuo-attentional exploration is terminated, lexical knowledge is updated. This final mechanism simulates either the reinforcement of the orthographic representation of a known word or the creation of a new lexical trace, both reflecting orthographic learning.

Therefore, overall, we have devised a model that visually explores a string stimulus, judging whether it is novel or not, with a unique exploration criterion based on the goal to obtain good perceptual representations of letters. At this point, our aim is thus to first characterize the visuo-attentional trajectories predicted by the model, and second, to assess whether these predictions match with eye movement patterns behaviorally observed during novel word orthographic learning.

3 Simulation of orthographic learning: the effect of repeated reading of novel words on eye movements

We now present simulation results from the *BRAID-Learn* model. We first illustrate the model's behavior on an example to detail how visuo-attentional exploration is performed and its consequences on letter identification and word processing. Then, we explore the model's behavior over successive exposures to a set of known and novel words. A new word representation was expected to be created for each novel word that was recognized as such. We were specially interested in how the strengthening across exposures of the newly created word representations would affect the number and duration of visuo-attentional captures. As gaze position and the focus of visual attention were aligned in the model, the measure of the number of visuo-attentional captures can be compared with the number of fixations in behavioral experiments, and the duration of visuo-attentional captures to fixation duration. As known words had a fully specified lexical representation prior to the first exposure, a greater effect of the number of exposure on the two measures was expected for novel words than for known words. Last, to assess the model's plausibility, we checked whether its output behavior mimicked the pattern of eye movements reported for humans in similar conditions of orthographic learning.

3.1 Simulating orthographic learning: an illustrative example

First, we applied the *BRAID-Learn* model on a word already part of the known lexicon, the word *MENSONGE* (*LIE*), and second on a novel word to learn, *SCRODAIN* (pronounced / skrodɛ̃ /). In both cases, we analyzed simulation results both in terms of the output behavior, that is to say, the visuo-attentional displacements generated during exploration of the letter-string, and further, by showing how internal probability distributions evolved dynamically during the course of the simulation.

3.1.1 Applying BRAID-Learn to a known word

To illustrate orthographic learning on a known word, we re-used the same stimulus as when we illustrated the tasks of letter recognition, word recognition and lexical decision (see Fig. 3). However, here, instead of processing the stimulus with fixed central gaze and attention positions, we let the *BRAID-Learn* model select visuo-attentional parameters to optimize the accumulation of perceptual evidence over letters.

The simulation yielded two fixations for processing the word *MENSONGE*. The first one was dictated by default parameters of the *BRAID-Learn* model: whatever the word type, the first fixation for an 8-letter long stimulus is at position $g = \mu_A = 3$ (over the "N" of *MENSONGE*), with attentional dispersion $\sigma_A = 1.75$, and lasts 290 iterations. The second fixation, selected by optimizing the predicted perceptual information gain, was at position 7 (over the "G" of *MENSONGE*), with attentional dispersion $\sigma_A = 2.0$, and lasts 250 iterations.

At the end of the second fixation, the termination criterion was met and the model proceeded to orthographic learning. In the present case, the stimulus was a word, and correctly corresponded to the one recognized by the model, so that the lexical representation for word $W = MENSONGE$ was updated from the acquired perceptual representation over letters (note that the *BRAID-Learn* model performs this update irrespective of item type).

The time-course evolution of probability distributions over letters, over words, and over lexical membership during the simulation are shown in Fig. 7. We observe that the *BRAID-Learn* model had almost no effect on the dynamical evolution of word recognition (compare middle plots of Fig. 3 and Fig. 7) and lexical membership (compare bottom plots of Fig. 3 and Fig. 7), except for a slight increase in slope of probability curves at the beginning of the second fixation (iterations 290 to 310). This indicates that the selected fixation was slightly advantageous for word identification and lexical decision, as it slightly speeded up convergence toward high probability values.

For letter recognition, in contrast, the effect of the *BRAID-Learn* model was more drastic (compare top plots of Fig. 3 and Fig. 7). The first fixation mostly allowed identification of the letter directly under the fixation position (the "N" at position 3). In contrast, the second fixation, at position 7, almost boosted all remaining letters. Indeed, letters "N" and "G" at positions 6 and 7 were rapidly identified. Finally, the remaining letters, even far from fixation, also saw their probabilities ramp up and converge to high values, thanks to lexical influence, at this stage in full effect, and to the very high probability value for the word *MENSONGE* in lexical space.

3.1.2 Applying BRAID-Learn to a novel word

We then applied the *BRAID-Learn* model to an 8-letter non-word stimulus, the letter sequence *SCRODAIN*. For the first exposure, the simulation yielded 5 different fixations before the termination criterion was met; for the second exposure, 3 fixations were needed; for the third and subsequent exposures, 2 fixations were needed, in positions 3 then 7, exactly as in the previous example *MENSONGE*, in which the stimulus was a known word. Details about fixations for the first three exposures to stimulus *SCRODAIN* are shown in Fig. 8.

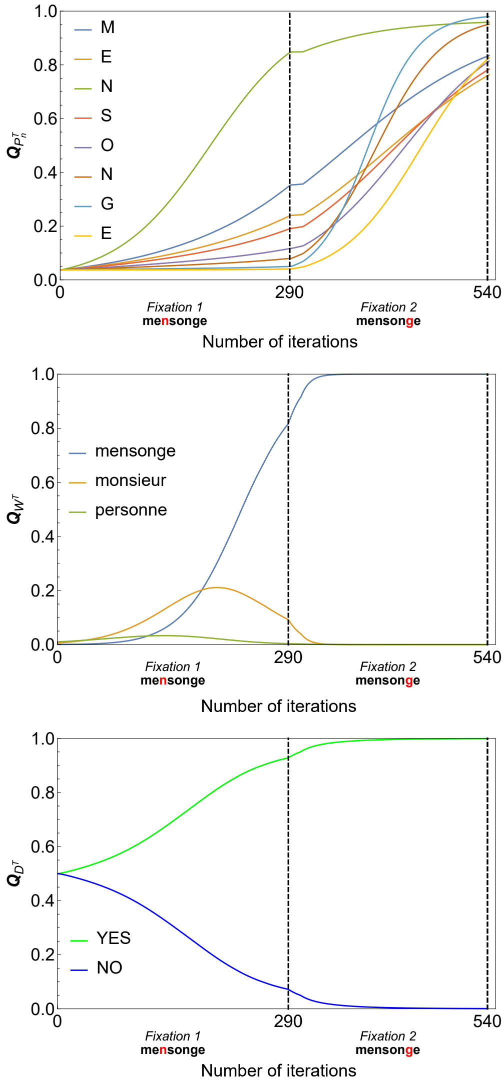

Fig. 7 Evolution of inference for $Q_{P_n^T}$ (top; Eq. (1)), Q_{W^T} (middle; Eq. (2)) and Q_{D^T} (bottom; Eq. (3)) as a function of simulated time (x-axis) for the 8-letter stimulus *MENSONGE*, with fixations computed by the *BRAID-Learn* model. Graphical representation is identical to the one of Fig. 3, with an added vertical, dashed line for delimiting different fixations.

Fig. 9 shows how Total Gain evolved as a function of exposures and fixations. We observe a stabilization of expected Total Gain after the third exposure, as the system converged towards a regime where stimulus *SCRODAIN*, having been already encountered three times, was associated with a lexical representation precise enough so that the stimulus was treated as a known word.

However, the first exposure appears to be different, with a Total Gain inferior to that of subsequent exposures. Recall that the Total Gain measure mostly captures the expected information gain during stimulus processing. During the first exposure, *SCRODAIN* was correctly identified as being a non-word, which, via the modulation of γ, drastically reduced the top-down transfer of information from the lexical submodel to the perceptual letter submodel. Consequently, during the first exposure, the only source of information about letter originated from sensory processing, contrary to subsequent exposures, where it originated both from sensory processing and lexical feedback. Information gain during first exposure was therefore smaller, overall, than for further exposures; the termination criterion based on information

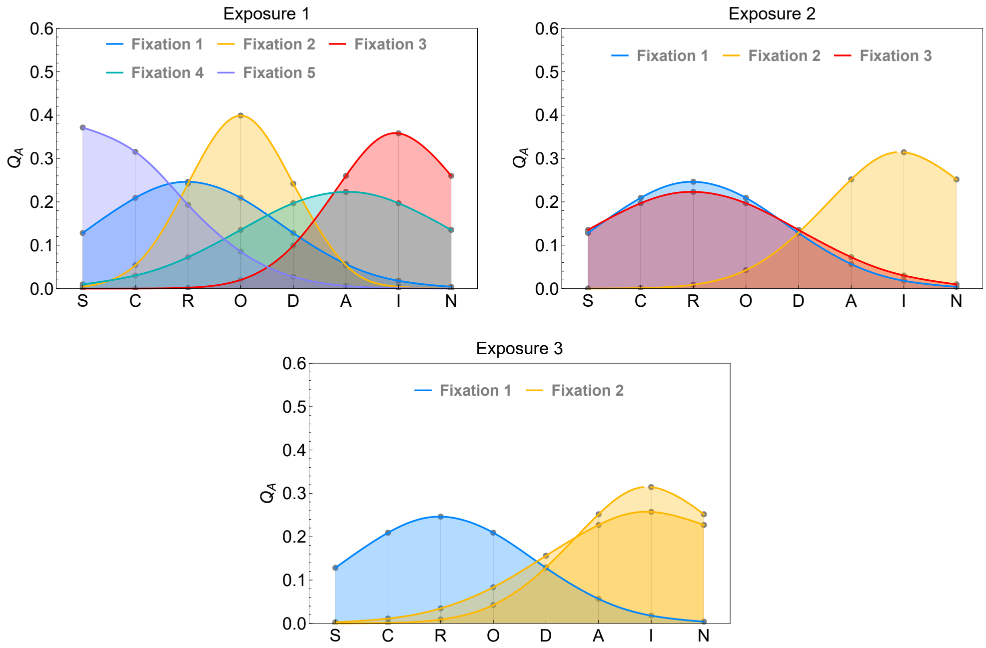

Fig. 8 Evolution of visuo-attentional parameters selected by the *BRAID-Learn* model during the first three exposures to the 8-letter stimulus non-word *SCRODAIN*. Each plot represents the probability value Q_A attributed to each position by the attentional model, following a Gaussian probability distribution. The mean value of each Gaussian distribution provides the selected position for attention focus $\mu_A^{T,f+1,e}$ and gaze position $g^{T,f+1,e}$: the first exposure (top left) yields 5 fixations (positions 3, 4, 7, 6 then 1); the second exposure (top right) yields 3 fixations (position 3 then position 7 then back to position 3); the third exposure yields 3 fixations (position 3 then 7 and 7 again), with the last two aggregated in our analyses, as they coincide in position. The standard-deviation of each Gaussian distribution provides the selected value for the dispersion of the visual attention distribution, $\sigma_A^{T,f+1,e}$.

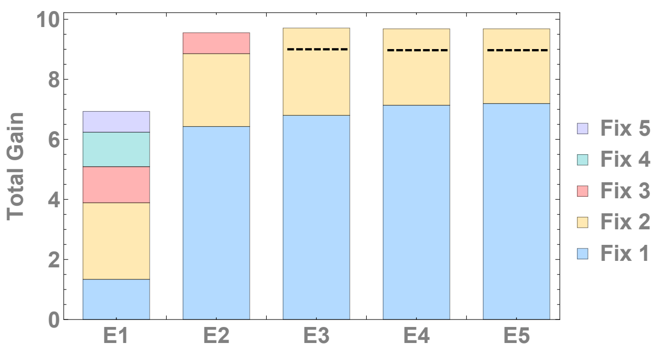

Fig. 9 Total Gain as a function of exposure number (noted E1 to E5) and fixation number (noted Fix 1 to Fix 5), during the processing of the 8-letter stimulus non-word *SCRODAIN*. The dashed horizontal line represents subsequent fixations that occur on the same spatial position, and are thus aggregated in following analyses.

accumulation speed was thus attained for higher values of remaining information. This explains how the first exposure had a smaller Total Gain value to reach before termination, compared to further exposures.

Fig. 10 shows the time-course evolution of probability distributions over letters, over words and over lexical membership during the first exposure to stimulus *SCRODAIN*. We observe that, when processing terminated, the stimulus was correctly recognized as a new word (the probability that D^T is false is high),

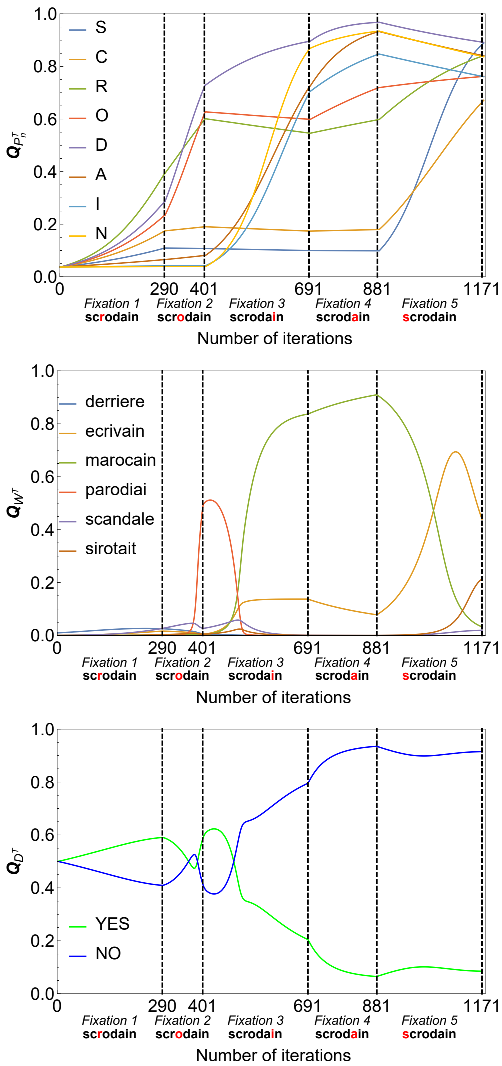

Fig. 10 Evolution of inference for $Q_{P_n^T}$ (top; Eq. (1)), Q_{W^T} (middle; Eq. (2)) and Q_{D^T} (bottom; Eq. (3)) as a function of simulated time (x-axis) for the first exposure to the 8-letter stimulus non-word *SCRODAIN*, with fixations computed by the *BRAID-Learn* model. Graphical representation is identical to the one of Fig. 7.

and all its letters were correctly identified (each probability distribution over letters, at each position, had a high value on the correct letter identity).

Since the stimulus was recognized as a new word, the probability distribution over words was switching between hypotheses, with no clear convergence to a single, winning hypothesis. This is the expected behavior, since, during first exposure to a novel word, the word space W does not contain a point corresponding to the stimulus. Instead, the most likely hypotheses in word space were close competitors to the stimulus, with the best one depending on processing stage, and more specifically, depending on current letter perception and fixation position. For example, consider iteration 401: few letters were well identified and gaze and attention were centered on the 4th position (the “O” of *SCRODAIN*). At this point, the most probable word was *PARODIAI* (*PARODIED*), which shares with *SCRODAIN* the “R”, “O” and “D”, which were the best perceived letters.

3.2 Simulation of visual processing during orthographic learning

Simulations were performed to, first, characterize the visuo-attentional trajectories predicted by the *BRAID-Learn* model, and, second, to assess how well predicted trajectories fit with the observed evolution of eye movement patterns during the orthographic learning of novel words. Only a few studies have reported the exposure-by-exposure evolution of eye movement patterns in conditions of incidental orthographic learning of novel words while reading (Ginestet et al., 2020; Joseph & Nation, 2018; Joseph et al., 2014; Pellicer-Sanchez, 2016). These studies consistently showed a decrease in reading times over exposures. The two studies that evaluated the effect of repeated exposures on both known words and novel words reported a decrease in reading times and number of fixations over exposures that was higher for novel words than for known words (Ginestet et al., 2020; Pellicer-Sanchez, 2016). The study of Ginestet et al. (2020) is singular in that it reported evidence on the evolution of eye movement patterns across exposures for items that were presented out of context. As the *BRAID-Learn* model only deals with isolated word processing, we assessed whether the model was able to simulate the effects reported for humans in the experimental study of Ginestet et al. (2020).

3.2.1 Material and method

Stimuli The set of items was the same as in Ginestet et al. (2020)'s study. It comprised 30 bisyllabic 8-letter novel words (among which was our previous example pseudoword, *SCRODAIN*) and 30 8-letter words (among which was our previous example word *MENSONGE*). Novel words were constructed from existing trigrams in French; they were graphotactically legal, none was homophone to a real word and none had any orthographic neighbor (i.e., words that differ from them by a single letter). The thirty words had no orthographic neighbors and were of medium frequency (per million, mean $f_W = 35.57$; SD $f_W = 18.86$). The list of items (novel and known words) is provided in Appendix A.

Method As participants of the behavioral experiment were adult French-speakers, the model was configured with lexical knowledge from the French lexicon Project (Ferrand et al., 2010), that is to say, its known words and frequency distribution were identified from that database of 38,840 French words. We first checked that all real words used in the experiment had a lexical entry in the model (and, of course, that the novel words did not). Then, the known words and novel words were presented five times to the *BRAID-Learn* model.

We first checked the model's capacity to recognize the input as either a known word or a novel word. Then, we assessed whether the model behavior across successive exposures exhibited the same main effects of item type and number of exposures, and the same item-type-by-exposure interaction as in the behavioral study. More specifically, novel word processing was expected to generate a higher number of fixations and longer fixation duration than the processing of known words. The two eye movement measures were expected to decrease as the number of exposures increased. Furthermore, the decrease of fixation number and fixation duration over exposures would be higher for the novel words. Last, as in Ginestet et al. (2020)'s and Pellicer-Sanchez (2016)'s study, the decrease of these two measures would be larger during the first exposures.

3.2.2 Simulated results

Overall, the model correctly processed 91.7% (55/60) of the items. All of the 30 control words were accurately recognized as known words and most novel words (25 out of 30) were accurately recognized as unknown during the first exposure, so that a new trace corresponding to each of the novel words was created; during subsequent encounters, each novel word was recognized as a known word and the recently created trace was strengthened in the word space W. For the remaining 5 novel words, the model incorrectly identified the stimulus as being a previously known word (final probability of lexical familiarity above .90), so that no new trace was created for this novel word. Instead, the most probable word, in all cases a close competitor of the stimulus in W (e.g., *CHANTANT* (*SINGING*) for *CHANQUET*), was chosen as the most likely hypothesis, yielding incorrect merging of the current perceptual trace with the lexical representation of the recognized word.

For the correctly processed items, we empirically observed that the simulated behavior differed between known and novel words. Processing was highly systematic for the known words, which were always processed in two fixations, located at Position 3 (set by calibration), then Position 7 (chosen by the entropy gain maximization mechanism). This highly systematic behavior did not follow from a predefined property of the *BRAID-Learn* model, but resulted from the entropy gain maximization principle.

In contrast, processing was far more variable for the novel words. Some novel words required five fixations at the first exposure, as in the above example for *SCRODAIN*, see Section 3.1.2). However, the number of fixations varied from three (e.g., for the novel word *PHACRAIT*), to six (e.g., for *PRIQUOIN*) at the first encounter. More importantly however, the number of required fixations systematically decreased for the novel words across exposures. In most cases, only two attentional fixations were predicted for the fifth exposure, which were located at position 3 then position 7, exactly as previously reported for words. Processing time (i.e., computed as the sum of all gaze durations on the input letter string) was shorter for known words than for novel words.

The statistical analyses were limited to the correctly processed items (i.e., 25 novel words and 30 words). We focused on the two measures of number of fixations and processing time, and on the item-type by exposure interaction, as in the experimental data. Results are presented in Fig. 11. An item-level analysis is presented in Fig. A.1 in Appendix B.

The simulated Number of Fixation and Processing Time measures were both analyzed by means of generalized linear mixed effects models (`glmer` function; R Core Team, 2018; RStudio version 1.3.1073). We used the Poisson family and the identity link for the analysis of number of fixations and the Gamma family and the identity link for processing times. Initially, a maximal random effects structure was specified including item random slope and intercept (Barr et al., 2013). While this full model converged for the analysis of processing times, it did not for the analysis of the number of fixations. Therefore, we followed the guidelines of Barr et al. (2013) and first removed correlations between random factors then random slopes, then random intercepts, to recover model convergence. Therefore, our analysis of the number of fixations ultimately amounts to using generalized linear model (`glm` function) instead.

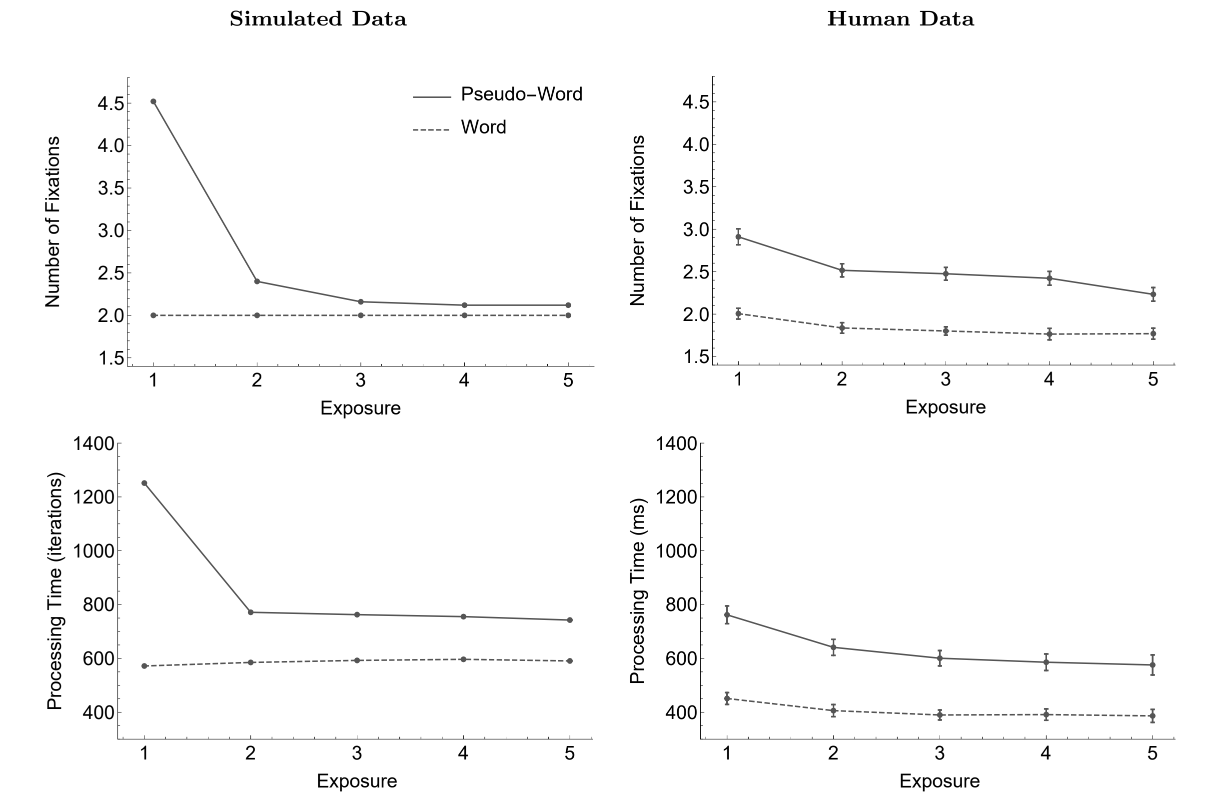

Fig. 11 Number of fixations (top plots) and processing time (bottom plots), as a function of exposures, from the reading aloud task in the experiment of Ginestet et al. (2020) (right plots) and simulated by the *BRAID-Learn* model (left plots), for known word stimuli (dashed lines) and novel word stimuli (solid lines).

In all models, contrasts were specified as 0.5/-0.5 or 2/1/0/-1/-2 when independent variables have, respectively, 2 or 5 modalities. All statistical models and simulated results are provided in Supplementary Material[1] (for a quick access, see the .html file from "Statistical_files" folder).

All statistical models included the number of exposures (from 1 to 5), item type (novel vs known words) and their interactions as fixed factors. Two post-hoc analyses were further conducted, first for local comparisons between the first and second exposures, and between the second and third exposures using similar models as previously described and, second, for comparisons of the two item type on the different measures (with the number of exposures as a fixed factor).

We first report statistical analyses concerning the number of fixations. Results showed main effects of item type ($\beta = -0.66,\ z = -3.57,\ p < .001$) and number of exposures ($\beta = 0.22,\ z = 3.37,\ p < .001$) on the number of fixations. As in the experimental data, the number of fixations across exposures decreased faster for novel words than for known words ($\beta = -0.44,\ z = -3.37,\ p < .001$). Post-hoc comparisons showed that this decrease mainly occurred between the first and the second exposure ($\beta = 2.12,\ z = 3.31,\ p < .001$; non significant interaction between the second and third exposure: $\beta = 0.24,\ z = 0.43,\ p = .669$).

We now turn to analyses of processing times. Results showed main effects of item type ($\beta = -265.98,\ t = -17.56,\ p < .001$) and number of exposures ($\beta = 40.87,\ t = 9.40,\ p < .001$). Processing time de-

[1] Open access availability for Supplementary Material files: https://osf.io/se645/?view_only=5d402ed3471f4492a4b12231f7ee7c09

creased more rapidly for novel words than for words ($\beta = -92.06,\ t = -10.72,\ p < .001$). Post-hoc comparisons showed that this interaction mainly occurred between the first and the second exposure ($\beta = 493.40,\ t = 14.85,\ p < .001$), with no statistically significant interaction between the second and third exposures ($\beta = 8.42,\ t = 0.33,\ p = .743$).

To summarize, as observed in human participants, the *BRAID-Learn* model successfully predicts different visuo-attentional trajectory characteristics for known words and novel words. The two main features of orthographic learning are reproduced in the simulations: a larger reduction in both processing time and number of fixations for novel words than for known words across the five exposures and a large decline between the first and the second exposure. Nevertheless, as shown in Fig. 11, there are some differences in magnitude between simulations and observations; in particular, for the first exposure, the number of fixations and processing time were far larger in the model than in experimental observations.

4 Discussion

The main contribution of the present study is the development and description of the *BRAID-Learn* model. The model makes original assumptions about the mechanisms allowing gradual extraction of visual information about letter identity during reading, how this information contributes to creating lexical orthographic knowledge and how each new exposition to the same word allows updating its lexical representation.

A strong postulate of *BRAID-Learn* is that visual attention is a core mechanism of orthographic learning. Visual attention was here implemented as a dynamic perceptual filter that allows selecting where information on letter identity should be extracted from the input word to optimize the speed of perceptual evidence accumulation. This makes orthographic learning possible and efficient.

Last, we have demonstrated through simulations that the model could successfully account for the overall shape of the evolution of eye movement patterns during orthographic learning. This is strong evidence in support of the model's assumptions, all the more that *BRAID-Learn* was neither specifically designed nor precisely configured or calibrated to account for eye movements while reading.

4.1 Theoretical contribution of the *BRAID-Learn* model

Our main contribution in the present paper is to make new assumptions about the mechanisms involved in the acquisition of orthographic knowledge and describe *BRAID-Learn*, the first computational model where the central focus is orthographic learning. Although the models of reading acquisition developed within the self-teaching framework (Pritchard et al., 2018; Ziegler et al., 2014) were designed to be able to enrich their orthographic knowledge through the acquisition of new orthographic representations, they did not implement the visuo-orthographic mechanisms involved in novel word orthographic learning. These models were derived from models of reading aloud, thus placing emphasis on phonological skills. In contrast, *BRAID-Learn* is the extension of a word recognition model. As a result, *BRAID-Learn* is explicit on the mechanisms of visual and visual attentional processing that are involved in the identification of the input word letter-string. However, the model does not include any of the phonological components usually postulated to account for reading aloud and reading acquisition. Thus, according to *BRAID-Learn*, orthographic learning is mainly conditional on the efficiency with which letters are identified within the novel word input string, while it is mainly conditional on successful phonological decoding according to self-teaching models.

These two views of orthographic learning are in no way contradictory. Quite the opposite: it is likely that both approaches shed light on two complementary facets of a complex process. There is strong evidence that successful phonological decoding contributes to orthographic learning, but the additional involvement of mechanisms of orthographic processing is largely acknowledged, even by the proponents of the self-teaching hypothesis (Castles & Nation, 2006; Castles et al., 2018; Pritchard et al., 2018). In the same way, *BRAID-Learn* describes the mechanisms of visuo-orthographic processing that are involved in orthographic learning, without precluding that additional factors, like phonological decoding or semantic knowledge, further contribute to the acquisition of new orthographic knowledge. Overall, *BRAID-Learn* sheds light on a facet of orthographic learning that was largely ignored by previous computational models. In this respect, *BRAID-Learn* paves the way for the development of a new generation of reading acquisition models that would combine both the visual and attentional mechanisms of orthographic processing and the mechanisms of phonological processing in a single framework.

It is further noteworthy that some of the assumptions of the *BRAID-Learn* model might be relevant to our conception of the reading system, in general. For example, the model postulates that familiarity judgment is assessed in an ongoing manner on input strings, and that the impact of top-down lexical knowledge on letter perception is modulated as a function of this lexical familiarity. This has a number of theoretical implications. First, if the pivotal role of the familiarity detector in orthographic learning is attested in future research, then this process should be considered as an integral part of the reading system, rather than being specific to the non-ecological experimental set-up of lexical decision tasks (Coltheart et al., 2001; Ginestet et al., 2019; McClelland & Rumelhart, 1981; Saghiran et al., 2020).

Second, the hypothesis that top-down lexical knowledge influences letter processing is not new. This feedback loop was mainly introduced to account for the word superiority effect, namely the fact that letters are more accurately recognized within words than when presented in isolation or within an unknown letter-string (Coltheart et al., 2001; McClelland & Rumelhart, 1981; Perry et al., 2007). However, alternatives to the interactive explanation of the word superiority effect have been proposed (Grainger & Jacobs, 1994; Paap et al., 1982) and a debate persists, to this day, about the relevance of such feedback loops during sensory processing (see, e.g., Magnuson et al. (2018) contra Norris et al. (2018)). Beyond the word superiority effect, the current findings suggest that top-down lexical influence is critical to account for the effects of orthographic learning on processing time and eye movements while reading. Independent evidence that lexical knowledge contributes to letter perception provides support to the interactive account of the word superiority effect.

Last, the online modulation of top-down lexical influence allowed configuring the model so that, from the same mathematical principle of perceptual evidence gain maximization, it would yield different oculomotor behaviors for known words and novel (or pseudo-) words. Evidence that different reading patterns can be generated for words and pseudo-words without any processing mechanism specific to the input item type is new evidence that should contribute to the debate between the dual route versus single route account of the reading system (Ans et al., 1998; Seidenberg, 2012; Seidenberg & McClelland, 1989).

A second postulate of the model of potential theoretical relevance is that visual attention appears as a core process of orthographic learning and word recognition. An overview of reading models shows that this is not consensual. In fact, many models of word recognition and reading aloud did not consider visual attention as part of the reading system (Coltheart et al., 2001; Davis, 2010; Gomez et al., 2008; McClelland & Rumelhart, 1981; Perry et al., 2007, 2010; Seidenberg & MacDonald, 1999; Whitney, 2001) and those that did were the exception (Ans et al., 1998; Ginestet et al., 2019; Mozer & Behrmann, 1990). This is all the more confusing that visual attention is described as an integral part of models of eye movement control (Engbert et al., 2002; Engbert et al., 2005; Reichle et al., 1999, 2003; Snell et al., 2018). Although the debate, there, focuses on whether attention is allocated to only one word at a time or to multiple words in parallel, eye movement control models do agree that word recognition is performed under the focus of attention. Evidence for an involvement of visual attention in orthographic learning might be additional evidence for reconsidering the role of visual attention in word recognition models.

4.2 The mechanisms of orthographic learning

In *BRAID-Learn*, orthographic learning is described as involving two mechanisms that affect letter identity processing, visual attention and lexical membership evaluation, along with a mechanism for orthographic memorization. As in Pritchard et al. (2018)'s self-teaching based model, the memorization process in *BRAID-Learn* varies depending on the familiarity of the input letter-string. Perceptual information on letter identity is either used to create a new orthographic representation if the input string has never been seen before or it is integrated with previous lexical knowledge if the input corresponds to an already seen word. A particular feature of the *BRAID-Learn* memorization process is that orthographic learning is not triggered by spoken-word recognition, so that new orthographic information can be learned without previous knowledge of corresponding phonological features.

In the model, we also assume that the top-down influence of word knowledge on letter perception is stronger when the stimulus appears to be a known word, than when it appears to be a novel word. At the first exposure with the novel word, lexical influence is decreased, so that most evidence that accumulates on letter identity at the perceptual level comes from bottom-up information. During orthographic learning, with successive exposures, the lexical representation of the novel word gets internalized, with probability distributions becoming less uncertain (i.e., their entropy decreases). The more the entropy of the newly created lexical representation decreases, the more it boosts letter identity information accumulation through top-down influence. Thus, modulation of lexical feedback depending on lexical familiarity is critical to avoid lexicalisation errors at the first exposure with a novel word and to successfully simulate the evolution of performance during orthographic learning. Although *BRAID-Learn* proposes a novel implementation of the familiarity detection mechanism and how it affects orthographic learning, Pritchard et al. (2018) also assumed that orthographic learning should be modulated depending on the visual familiarity of the input letter string.

The main originality of the *BRAID-Learn* model is to postulate that visual attention is at the core of orthographic learning. In *BRAID-Learn*, visual attention is conceived as a dynamic process that shifts over the input letter-string to try and allocate more attention to those letters that are more difficult to identify. In the original *BRAID* model (Phénix, 2018; Phénix et al., submitted), the parameters of the visuo-attentional distribution were fixed and their default values used in all simulations, independently of the input word characteristics (except for long words, see Ginestet et al. (2019)). In contrast, *BRAID-Learn* uses dynamic visuo-attentional parameters, and their values are computed online during processing with the purpose to optimize the gain of information on letter identity at each fixation. Although rarely applied to the modeling of reading (Bernard et al., 2008; Legge et al., 2002; Legge et al., 1997; Salvucci, 2001), the assumption that visual processing aims at optimizing the speed of perceptual evidence accumulation is common in a wide variety of domains, including computational modeling of oculomotor behavior during natural scene visual perception (Lee & Stella, 2000; Raj et al., 2005), visual search modeling (Colas et al., 2009; Friston et al., 2012; Najemnik & Geisler, 2005; Navalpakkam et al., 2010) and the modeling of visual exploration of objects (shape-matching, Renninger et al., 2007). In the Mr. Chips model of text reading (Legge et al., 2002; Legge et al., 1997), it is assumed that saccade length is selected to minimize uncertainty about the fixated word and that refixations occur until the fixated word is perfectly identified. Similarly, in their simulation of differences in reading strategies in normal readers and central scotoma patients during word recognition, Bernard et al. (2008) assumed that word recognition occurs through an optimal reading strategy, in which gaze fixation locations are selected in order to maximize information gain about letters. However, contrary to the *BRAID-Learn* model, these models did not represent visual attention.

BRAID-Learn, for the first time, provides a description of the dynamics of visual attention for efficient letter perception and shows that flexibility in the visuo-attentional distribution over time makes orthographic learning possible and efficient. The key role of visual attention in orthographic learning that is predicted by *BRAID-Learn* contrasts with previous accounts by self-teaching based models. However, as noted in previous sections, these models made the simplifying assumption that complete information on the whole input letter string was available in a one-shot manner, from the first exposure with the novel word. In *BRAID-Learn*, orthographic learning is gradual and visuo-attentional captures over the letter string help gather information efficiently on letter identity; this improves perceptual evidence accumulation and stabilizes orthographic representations after a few exposures.

Whether visual attention affects word recognition and orthographic learning is a controversial issue. Despite behavioral evidence that visual attention is involved in printed word recognition (Besner et al., 2016; Lachter et al., 2004; Risko et al., 2010; Waechter et al., 2011) and in reading acquisition (Bosse & Valdois, 2009; Valdois et al., 2019), most computational models of word recognition and the self-teaching models do not incorporate any visual attentional mechanism (Coltheart et al., 2001; Perry et al., 2007, 2010; Pritchard et al., 2018; Seidenberg & McClelland, 1989; Ziegler et al., 2014). It is also quite puzzling that the critical role attributed to visual attention in the perceptual learning of new orthographic units, from letters to words, by Laberge and Samuels (1974) in the first model of reading acquisition ever proposed was largely ignored by subsequent modelling attempts. In the same way, very little systematic research has been directed to the role of visual attention in orthographic learning. However, some recent behavioral evidence provides support to the *BRAID-Learn* predictions. Investigation of incidental learning while reading in adult skilled readers suggested that orthographic learning was more efficient in individuals who had a higher visual attention span (Ginestet et al., 2020). Higher visual attention capacity also accounted for better orthographic learning in typical children of opaque languages (Marinelli et al., 2020). These findings suggest that visual attention would modulate the orthographic learning behavior in humans. Assuming that orthographic learning is a foundation for both fast word recognition and word spelling, visual attention should further modulate these skills. This seems to be the case. Indeed, there is growing evidence that visual attention is a concurrent and longitudinal predictor of word reading fluency (Bosse & Valdois, 2009; Chan & Yeung, 2020; Valdois et al., 2019; van den Boer & de Jong, 2018) and word spelling acquisition (Niolaki et al., 2020; Valdois et al., submitted; van den Boer et al., 2015) and that individuals with reduced visual attention capacity are slow readers and poor spellers (Bosse et al., 2007; Chen et al., 2019; Valdois et al., 2011; Valdois et al., 2021; Zoubrinetzky et al., 2014). Evidence that *BRAID-Learn* can account for the evolution of eye movement patterns when repeatedly confronted to the same input strings is further evidence in support of its theoretical assumptions.

4.3 Prediction of the evolution of eye movement patterns during orthographic learning

Because it describes the dynamics of visual attention for letter identification within the input string and the evolution of lexical influence during processing, *BRAID-Learn* was expected to account for at least some of the changes that characterize eye movement patterns during the course of orthographic learning. Simulation results suggest that the model is rather efficient in doing so. First, the model predicts a larger number of fixations and longer processing time for novel words than for known word at the first encounter, well in line with the differential oculomotor patterns reported in humans when confronted to known words versus pseudowords or to words that drastically differ in frequency (Chaffin et al., 2001; Lowell & Morris, 2014; Rau et al., 2015; Wochna & Juhasz, 2013). Second, a decrease in number of fixations and processing time is predicted across exposures for the novel words, which again matches the oculomotor pattern changes reported in humans following multiple exposures to the same pseudowords (Gerbier et al., 2015, 2018; Joseph & Nation, 2018; Joseph et al., 2014). Last, in line with the behavioral data showing strong variation of oculomotor patterns between the first and the second exposure (Ginestet et al., 2020; Pellicer-Sanchez, 2016) and robust orthographic learning, after a single exposure (Bowey & Muller, 2005; Nation et al., 2007; Share, 2004), the model predicts a sharp decrease of the number of fixations and processing time, as early as the second exposure.

BRAID-Learn provides an account on the way the mechanisms of orthographic learning may affect oculomotor behavior. During the first exposure to a novel word, visual attention moves over the input string to maximize information gain about letters while top-down lexical influence is decreased. High uncertainty on the identity of the letters within the novel word increases the probability of attention captures. The model is also more prone to focus attention over a subset of letters when identification is difficult, which is done to the detriment of the other letters' identification, and again favors subsequent attention captures, and thus refixations. This translates in a larger number of fixations and longer total processing time for novel words than for known words, since only the latter benefit from top-down lexical influence that boosts letter identification from the first encounter. However, top-down lexical information becomes effective from the second exposure to the same novel word. Information extracted on letter identity and memorized during the first exposure can then be used to speed-up letter processing. This positive effect of top-down influence from the newly acquired word orthographic representation results in large decreases of both number of fixations and processing time at the second exposure. Further exposures lead to further improvements of the orthographic representation of the word being acquired and variations in the strength of lexical feedback result in more gradual changes in eye movements. Overall, the model behavior is well in line with the empirical findings reported in humans, namely that the first exposure to a novel word is more critical than later ones for orthographic learning and that the orthographic representations of new words are stabilized, after only a small, single-digit number of exposures (Ginestet et al., 2020; Nation et al., 2007; Pellicer-Sanchez, 2016; Share, 2004).

Nevertheless, adjusting the model parameters could improve the model's predictions fit to data. In particular, the large number of fixations and long processing time generated at the first exposure with a novel word are rather unrealistic. This suggests that the termination criterion, based on the possible remaining entropy gain value, was underestimated. To recall, the corresponding threshold value was set to 1 nat of information gain for the whole word; this value was set arbitrarily, without calibration from empirical or experimental data. We observed that this arbitrary value yielded a "conservative" behavior, with the model terminating with "very certain" perceptual information about all letters. An increase of this value would stop visuo-attentional exploration faster, and thus reduce the number of fixations and processing time, even at the first encounter with the novel word. Further, increasing this threshold would yield a more gradual evolution of information gain over time, leading the oculomotor patterns of word and pseudo-word processing to converge later on, more in line with the human data. However, in a more complete model, the interaction between orthographic and phonological processing could affect the dynamics of visuo-attentional exploration; furthermore, the model currently does not take into account, for instance, any time interval for performing saccades; therefore, we consider that a proper calibration of the model's parameters to fit behavioral data is premature at this stage. A proper calibration of the model parameters to data would also entail collecting experimental data about visual exploration during orthographic learning of words of varying characteristics (such as, e.g., length, frequency, etc.); such data are currently not available.

The simulations were carried out to assess to what extent a model limited to visual orthographic processing would account for the features of eye movement patterns reported in humans during orthographic learning. The good qualitative account of eye movement pattern changes across exposures suggests that the two mechanisms of visual attention and lexical feedback postulated by the model are critical factors for orthographic learning in humans, and may be also in animals (Grainger et al., 2012; Scarf et al., 2016). It is noteworthy that *BRAID-Learn* was not specifically developed to account for eye movement patterns while reading. As such, *BRAID-Learn*, for the first time, offers a unified account of three aspects of the reading system that are typically modelled independently, namely word recognition, eye movement control and orthographic learning. Visual attention appears critical for an integrated account of these different dimensions of the reading process.

Acknowledgments

This work has been supported by a French Ministry of Research (MESR) grant to EG and the French government as part of the e-FRAN "FLUENCE" project funded by the "Investissement d'Avenir" program handled by the "Caisse des Dépôts et Consignations".

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Open Practices Statement

Simulated results and statistical files, including the R script and a direct online access to statistical results (html file), are available by following this link https://osf.io/se645/?view_only=5d402ed3471f4492a4b12231f7ee7c09.

Appendix A: List of items

Lists of Items used in simulations

Pseudo-words : broufand, chaiquau, deinrint, faingion, gouciont, nauplois, ploitart, speirain, quinsard, tramoint, ceitteau, chanquet, coirtint, drottont, flommais, glounein, priquoin, quarlant, siampoie, trimpond, bussiond, cherrein, ciercard, claffand, fentroit, phacrait, prinnant, tauppart, scrodain, trancare.
Control words : uniforme, portrait, enceinte, mouchoir, surprise, complice, scandale, chanteur, immeuble, avantage, physique, revanche, horrible, boutique, sensible, fauteuil, chocolat, mensonge, solution, voyageur, prochain, grandeur, nocturne, lointain, religion, empereur, division, quartier, province, jugement.

Appendix B: Item-level simulations

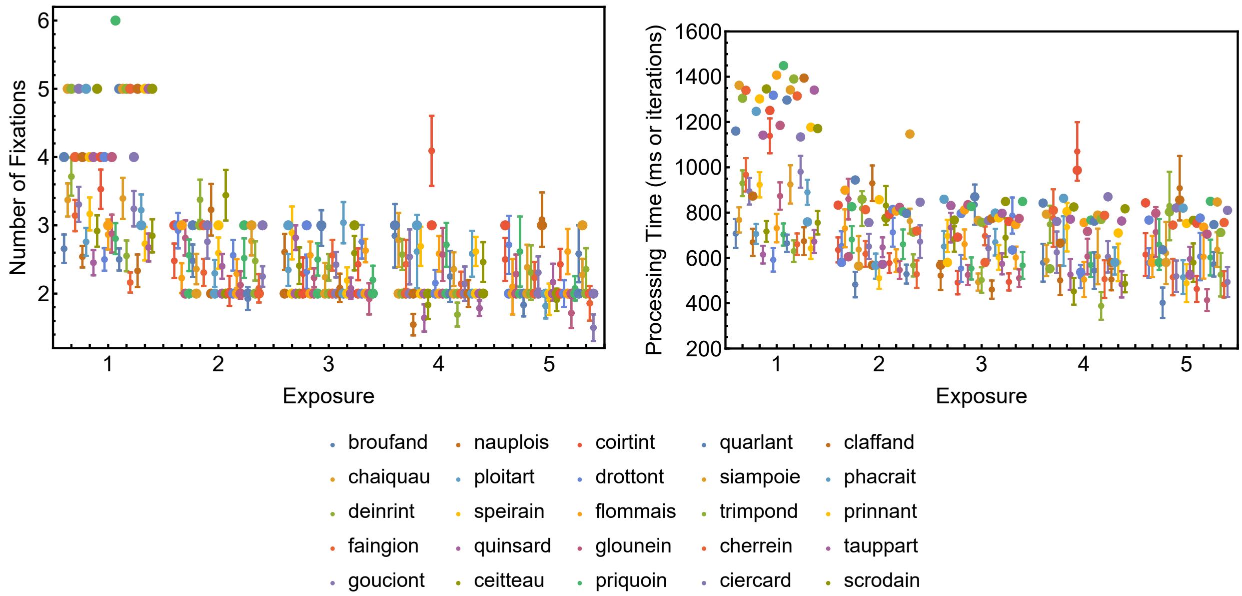

Fig. A.1 Number of Fixations (left plot) and Processing Time (right plot) reported as a function of exposures and for each novel word learned by the *BRAID-Learn* model. Smaller dots and their vertical bars (standard errors) represent behavioral data; larger dots represent simulated results.

References

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*(4), 678–723. https://doi.org/10.1037/0033-295X.105.4.678-723

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bernard, J.-B., del Prado Martin, F. M., Montagnini, A., & Castet, E. (2008). A model of optimal oculomotor strategies in reading for normal and damaged visual fields. *Deuxième conférence française de Neurosciences Computationnelles, "Neurocomp08"*.

Besner, D., Risko, E. F., Stolz, J. A., White, D., Reynolds, M., O'Malley, S., & Robidoux, S. (2016). Varieties of attention: Their roles in visual word identification. *Current Directions in Psychological Science*, *25*(3), 162–168. https://doi.org/10.1177/0963721416639351

Bessière, P., Laugier, C., & Siegwart, R. (Eds.). (2008). *Probabilistic reasoning and decision making in sensory-motor systems* (Vol. 46). Springer. https://doi.org/10.1007/978-3-540-79007-5

Bosse, M.-L. (2005). De la relation entre acquisition de l'orthographe lexicale et traitement visuo-attentionnel chez l'enfant. *Rééducation orthophonique*, *222*, 9–30.

Bosse, M.-L., Chaves, N., Largy, P., & Valdois, S. (2015). Orthographic learning during reading: The role of whole-word visual processing. *Journal of Research in Reading*, *38*(2), 141–158. https://doi.org/10.1111/j.1467-9817.2012.01551.x

Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*(2), 198–230. https://doi.org/10.1016/j.cognition.2006.05.009

Bosse, M.-L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: A cross-sectional study. *Journal of Research in Reading*, *32*(2), 230–253. https://doi.org/10.1111/j.1467-9817.2008.01387.x

Bowey, J., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders' silent reading: A critical test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *92*(2), 203–219. https://doi.org/10.1016/j.jecp.2005.06.005

Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 151–179). Psychology Press.

Castles, A., & Nation, K. (2008). Learning to be a good orthographic reader. *Journal of Research in Reading*, *31*(1), 1–7. https://doi.org/10.1111/j.1467-9817.2007.00367.x

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Chaffin, R., Morris, R., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 225–235. https://doi.org/10.1037/0278-7393.27.1.225

Chan, K. S.-C., & Yeung, P.-S. (2020). Prediction of chinese reading fluency by verbal and non-verbal visual attention span measures. *Frontiers in Psychology*, (10), 3049. https://doi.org/10.3389/fpsyg.2019.03049

Chaves, N., Totereau, C., & Bosse, M.-L. (2012). Acquérir l'orthographe lexicale: Quand savoir lire ne suffit pas. *ANAE – Approche Neuropsychologique des Apprentissages chez L'Enfant*, *118*, 271–279.

Chen, N. T., Zheng, M., & Ho, C. S.-H. (2019). Examining the visual attention span hypothesis in chinese developmental dyslexia. *Reading and Writing*, *32*, 639–662. https://doi.org/10.1007/s11145-018-9882-1

Colas, F., Flacher, F., Tanner, T., Bessière, P., & Girard, B. (2009). Bayesian models of eye movement selection with retinotopic maps. *Biological Cybernetics*, *100*(3), 203–214. https://doi.org/10.1007/s00422-009-0292-y

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology*, *95*, 56–77. https://doi.org/10.1016/j.jecp.2006.03.008

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing: An interdisciplinary Journal*, *14*, 549–568. https://doi.org/10.1023/A:1011100226798

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*(3), 713. https://doi.org/10.1037/a0019738

de Jong, P. F., Bitter, D. J., van Setten, M., & Marinus, E. (2009). Does phonological recoding occur during silent reading and is it necessary for orthographic learning? *Journal of Experimental Child Psychology*, *104*(3), 267–282. https://doi.org/10.1016/j.jecp.2009.06.002

Dubois, M., Kyllingsbæk, S., Prado, C., Musca, S. C., Peiffer, E., Lassus-Sangosse, D., & Valdois, S. (2010). Fractionating the multi-character processing deficit in developmental dyslexia: Evidence from two case studies. *Cortex*, *46*, 717–738. https://doi.org/10.1016/j.cortex.2009.11.002

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, *42*(5), 621–636. https://doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). Swift: A dynamical model of saccade generation during reading. *Psychological review*, *112*(4), 777. https://doi.org/10.1037/0033-295X.112.4.777

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. https://doi.org/10.3758/BRM.42.2.488

Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in psychology*, *3*, 151. https://doi.org/10.3389/fpsyg.2012.00151

Gerbier, E., Bailly, G., & Bosse, M.-L. (2015). Using Karaoke to enhance reading while listening: Impact on word memorization and eye movements. *Speech and Language Technology for Education (SLaTE)*, 59–64.

Gerbier, E., Bailly, G., & Bosse, M.-L. (2018). Audio-visual synchronization in reading while listening to texts: Effects on visual behavior and verbal learning. *Computer Speech and Language*, *47*, 79–92. https://doi.org/10.1016/j.csl.2017.07.003

Geyer, L. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics*, *22*(5), 487–490. https://doi.org/10.3758/BF03199515

Gilet, E., Diard, J., & Bessière, P. (2011). Bayesian action-perception computational model: Interaction of production and recognition of cursive letters. *PLoS ONE*, *6*(6), e20387. https://doi.org/10.1371/journal.pone.0020387

Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical decision: The role of visual attention. *Vision Research*, *159*, 10–20. https://doi.org/10.1016/j.visres.2019.03.003

Ginestet, E., Valdois, S., Diard, J., & Bosse, M.-L. (2020). Orthographic learning of novel words in adults: Effects of exposure and visual attention on eye movements. *Journal of Cognitive Psychology*, *32*(8), 785–804. https://doi.org/10.1080/20445911.2020.1823987

Gomez, P., Ratcliff, R., & Perea, M. (2008). The Overlap model: A model of letter position coding. *Psychological Review*, *115*(3), 577–601. https://doi.org/10.1037/a0012667

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in Baboons (*Papio Papio*). *Science*, *336*(6078), 245–248. https://doi.org/10.1126/science.1218152

Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1158. https://doi.org/10.1037/0096-1523.20.6.1158

Howard, D., & Best, W. (1996). Developmental phonological dyslexia: Real word reading can be completely normal. *Cognitive Neuropsychology*, *13*(6), 887–934. https://doi.org/10.1080/026432996381854

Inserm. (2007). Dyslexie, dysorthographie, dyscalculie. Bilan des données scientifiques. *Paris: Les éditions Inserm*, *15*, 842p–Expertise collective.

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Joseph, H., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238–248. https://doi.org/10.1016/j.cognition.2014.06.015

Kyte, C. S., & Johnson, C. J. (2009). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology*, *93*(2), 166–185. https://doi.org/10.1016/j.jecp.2005.09.003

Laberge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(10), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): Still no identification without attention. *Psychological Review*, *111*(4), 880–913. https://doi.org/10.1037/0033-295X.111.4.880

Lee, T. S., & Stella, X. Y. (2000). An information-theoretic framework for understanding saccadic eye movements. *Advances in neural information processing systems*, 834–840.

Legge, G. E., Hooven, T. A., Klitz, T. S., Mansfield, J. S., & Tjan, B. S. (2002). Mr. Chips 2002: New insights from an ideal-observer model of reading. *Vision research*, *42*(18), 2219–2234. https://doi.org/10.1016/S0042-6989(02)00131-1

Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological review*, *104*(3), 524. https://doi.org/10.1037/0033-295X.104.3.524

Lobier, M., Dubois, M., & Valdois, S. (2013). The role of visual processing speed in reading speed development. *PLoS ONE*, *8*(4), e58097. https://doi.org/10.1371/journal.pone.0058097

Lowell, R., & Morris, R. K. (2014). Word length effects on novel words: Evidence from eye movements. *Attention, Perception, & Psychophysics*, *76*(1), 179–189. https://doi.org/10.3758/s13414-013-0556-4

Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. D. (2018). Interaction in spoken word recognition models: Feedback helps. *Frontiers in Psychology*, *9*, 369. https://doi.org/10.3389/fpsyg.2018.00369

Marinelli, C. V., Zoccolotti, P., & Romani, C. (2020). The ability to learn new written words is modulated by language orthographic consistency. *PloS one*, *15*(2), e0228129. https://doi.org/10.1371/journal.pone.0228129

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Mozer, M. C., & Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, *2*(2), 96–123. https://doi.org/10.1162/jocn.1990.2.2.96

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387. https://doi.org/10.1038/nature03390

Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning to read English: Effects of exposure, durability, and context. *Journal of Experimental Child Psychology*, *96*(1), 71–84. https://doi.org/10.1016/j.jecp.2006.06.004

Nation, K., & Castles, A. (2017). Putting the learning into orthographic learning. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of reading development* (pp. 147–168). John Benjamins Publishing Company.

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences*, *107*(11), 5232–5237. https://doi.org/10.1073/pnas.0911972107

Niolaki, G. Z., Vousden, J., Terzopoulos, A. R., Taylor, L. M., Hall, S., & Masterson, J. (2020). Predictors of single word spelling in english speaking children: A cross-sectional study. *Journal of Research in Reading*, *43*(4), 577–596. https://doi.org/10.1111/1467-9817.12330

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal bayesian decision process. *Psychological Review*, *113*(2), 327–357. https://doi.org/10.1037/0033-295X.113.2.327

Norris, D., McQueen, J., & Cutler, A. (2018). Commentary on "interaction in spoken word recognition models". *Frontiers in Psychology*, *9*, 1568. https://doi.org/10.3389/fpsyg.2018.01568

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, *89*(5), 573–594. https://doi.org/10.1037/0033-295X.89.5.573

Pagan, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing and retrieval effects in adults. *Cognitive Science*, *43*. https://doi.org/10.1111/cogs.12705

Pellicer-Sanchez, A. (2016). Incidental l2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, *38*, 97–130. https://doi.org/10.1017/S0272263115000224

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*(2), 106–51. https://doi.org/10.1016/j.cogpsych.2010.04.001

Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale computational models. *Psychological science*, *30*(3), 386–395. https://doi.org/10.1177/0956797618823540

Phénix, T. (2018). *Modélisation bayésienne algorithmique de la reconnaissance visuelle de mots et de l'attention visuelle* (Doctoral dissertation). Univ. Grenoble Alpes.

Phénix, T., Ginestet, E., Valdois, S., & Diard, J. (submitted). Visual attention matters during word recognition: A Bayesian modeling approach. —.

Phénix, T., Valdois, S., & Diard, J. (2018). Reconciling opposite neighborhood frequency effects in lexical decision: Evidence from a novel probabilistic model of visual word recognition. In T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp. 2238–2243). Cognitive Science Society.

Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A computational model of the self-teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive Science*, 1–49. https://doi.org/10.1111/cogs.12571

R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Raj, R., Geisler, W. S., Frazor, R. A., & Bovik, A. C. (2005). Contrast statistics for foveated visual systems: Fixation selection by minimizing contrast entropy. *JOSA A*, *22*(10), 2039–2049. https://doi.org/10.1364/JOSAA.22.002039

Rajalingham, R., Kar, K., Sanghavi, S., Dehaene, S., & DiCarlo, J. J. (2020). The inferior temporal cortex is a potential cortical precursor of orthographic processing in untrained monkeys. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-17714-3

Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye movement behavior: German and English children and adults process the same words differently. *Journal of Experimental Child Psychology*, *130*, 92–105. https://doi.org/10.1016/j.jecp.2014.09.012

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision research*, *39*(26), 4403–4411. https://doi.org/10.1016/S0042-6989(99)00152-2

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, *26*(4), 445–476. https://doi.org/10.1017/S0140525X03000104

Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? eye movements reduce local uncertainty. *Journal of vision*, *7*(3), 6–6. https://doi.org/10.1167/7.3.6

Risko, E. F., Stolz, J. A., & Besner, D. (2010). Spatial attention modulates feature crosstalk in visual word processing. *Attention, Perception and Psychophysics*, *72*(4), 989–998. https://doi.org/10.3758/APP.72.4.989

Romani, C., Di Betta, A. M., Tsouknida, E., & Olson, A. (2008). Lexical and nonlexical processing in developmental dyslexia: A case for different resources and different impairments. *Cognitive neuropsychology*, *25*(6), 798–830. https://doi.org/10.1080/02643290802347183

Romani, C., Ward, J., & Olson, A. (1999). Developmental surface dysgraphia: What is the underlying cognitive impairment? *The Quarterly Journal of Experimental Psychology Section A*, *52*(1), 97–128. https://doi.org/10.1080/713755804

Saghiran, A., Valdois, S., & Diard, J. (2020). Simulating length and frequency effects across multiple tasks with the bayesian braid-phon model. In S. Denison, M. Mack., Y. Xu, & B. Armstrong (Eds.), *Proceedings of the 42nd annual conference of the cognitive science society*. Cognitive Science Society.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, *1*(4), 201–220. https://doi.org/10.1016/S1389-0417(00)00015-2

Scarf, D., Boy, K., Reinert, A. U., Devine, J., Güntürkün, O., & Colombo, M. (2016). Orthographic processing in pigeons (columba livia). *Proceedings of the National Academy of Sciences*, *113*(40), 11272–11276. https://doi.org/10.1073/pnas.1607870113

Seidenberg, M. S. (2012). Computational models of reading: Connectionist and dual-route approaches. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *Cambridge handbook of psycholinguistics* (pp. 186–203). Cambridge University Press.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive science*, *23*(4), 569–588. https://doi.org/10.1207/s15516709cog2304_8

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*(4), 267–298. https://doi.org/10.1016/j.jecp.2004.01.001

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). Ob1-reader: A model of word recognition and eye movements in text reading. *Psychological review*, *125*(6), 969. https://doi.org/10.1037/rev0000119

Tucker, R., Castles, A., Laroche, A., & Deacon, H. (2016). The nature of orthographic learning in self-teaching: Testing the extent of transfer. *Journal of Experimental Child Psychology*, *145*, 79–94. https://doi.org/10.1016/j.jecp.2015.12.007

Valdois, S., Bidet-Ildei, C., Lassus-Sangosse, D., Reilhac, C., N'guyen-Morel, M.-A., Guinet, E., & Orliaguet, J.-P. (2011). A visual processing but no phonological disorder in a child with mixed dyslexia. *Cortex*, *47*(10), 1197–1218. https://doi.org/10.1016/j.cortex.2011.05.011

Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing: An Interdisciplinary Journal*, *16*, 541–572. https://doi.org/10.1023/A:1025501406971

Valdois, S., Phénix, T., Fort, M., & Diard, J. (submitted). Atypical viewing position effect in developmental dyslexia : A behavioral and modeling exploration. —.

Valdois, S., Reilhac, C., Ginestet, E., & Line Bosse, M. (2021). Varieties of cognitive profiles in poor readers: Evidence for a vas-impaired subtype. *Journal of Learning Disabilities*, *54*(3), 221–233. https://doi.org/10.1177/0022219420961332

Valdois, S., Roulin, J.-L., & Bosse, M.-L. (2019). Visual attention modulates reading acquisition. *Vision Research*, *165*(1), 152–161. https://doi.org/10.1016/j.visres.2019.10.011

van den Boer, M., van Bergen, E., & de Jong, P. F. (2015). The specific relation of visual attention span with reading and spelling in Dutch. *Learning and individual differences*, *39*, 141–149. https://doi.org/10.1016/j.lindif.2015.03.017

van den Boer, M., & de Jong, P. F. (2018). Stability of visual attention span performance and its relation with reading over time. *Scientific Studies of Reading*, *22*(5), 434–441. https://doi.org/10.1080/10888438.2018.1472266

van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2013). Modeling the length effect: Specifying the relation with visual and phonological correlates of reading. *Scientific Studies of Reading*, *17*(4), 243–256. https://doi.org/10.1080/10888438.2012.683222

Waechter, S., Besner, D., & Stolz, J. A. (2011). Basic processes in reading: Spatial attention as a necessary preliminary to orthographic and semantic processing. *Visual Cognition*, *19*(2), 171–202. https://doi.org/10.1080/13506285.2010.517228

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, *8*(2), 221–43. https://doi.org/10.3758/BF03196158

Wochna, K. L., & Juhasz, B. J. (2013). Context length and reading novel words: An eye-movement investigation. *British Journal of Psychology*, *104*(3), 347–363. https://doi.org/10.1111/j.2044-8295.2012.02127.x

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. https://doi.org/10.1098/rstb.2012.0397

Zoubrinetzky, R., Bielle, F., & Valdois, S. (2014). New insights on developmental dyslexia subtypes: Heterogeneity of mixed reading profiles. *PLoS ONE*, *9*(6), e99337. https://doi.org/10.1371/journal.pone.0099337