



HAL
open science

Probabilistic modeling of orthographic learning based on visuo-attentional dynamics

Emilie Ginestet, Sylviane Valdois, Julien Diard

► To cite this version:

Emilie Ginestet, Sylviane Valdois, Julien Diard. Probabilistic modeling of orthographic learning based on visuo-attentional dynamics. *Psychonomic Bulletin and Review*, 2022, 10.3758/s13423-021-02042-4. hal-03624916

HAL Id: hal-03624916

<https://hal.univ-grenoble-alpes.fr/hal-03624916>

Submitted on 6 Oct 2022

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

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

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

39 orthographic knowledge, beyond phonological skills (Cunningham et al., 2001; see Castles and Nation,
40 2006 for a review).

41 Further evidence against phonological processing as the unique basis of orthographic learning comes
42 from developmental dyslexia. On the one hand, prototypical patterns of phonological dyslexia have been
43 observed in patients who demonstrate fully developed word-specific orthographic knowledge despite ma-
44 jor phonological deficit (Howard & Best, 1996). On the other hand, there are cases of surface dyslexia
45 who show major deficits of irregular-word reading and spelling despite normal phonological skills (Inserm,
46 2007; Romani et al., 2008; Romani et al., 1999; Valdois et al., 2003). This suggests that very poor phono-
47 logical decoding skills do not necessarily prevent orthographic learning and that having good phonological
48 decoding skills does not guarantee normal development of lexical orthographic knowledge. Of particular
49 interest for the present purpose, search for the cognitive deficits associated with developmental surface
50 dyslexia revealed that a selective orthographic deficit was associated with a deficit of the simultaneous
51 processing of distinct visual elements, dubbed the visual attention (VA) span deficit (Bosse, 2005; Dubois
52 et al., 2010; Valdois et al., 2003). Further evidence that VA span more specifically relates to reading sub-
53 skills that reflect word-specific orthographic knowledge – like irregular word reading (Bosse & Valdois,
54 2009), reading speed (Lobier et al., 2013; van den Boer et al., 2015; van den Boer & de Jong, 2018) or the
55 length effect in word reading (van den Boer et al., 2013) – supports a potential contribution of VA span
56 to word-specific orthographic knowledge acquisition. More direct evidence comes from studies showing a
57 link between VA span and spelling acquisition (Niolaki et al., 2020; van den Boer et al., 2015) and from
58 studies showing that VA span modulates novel word orthographic learning (Bosse et al., 2015; Chaves
59 et al., 2012; Ginestet et al., 2020; Marinelli et al., 2020). Without minimizing the role of phonological
60 skills in orthographic acquisition, these findings suggest that visual factors independently contribute to
61 the development of word-specific orthographic knowledge. Data from animal studies further suggest that
62 the contribution of visual processing skills to orthographic knowledge acquisition may have been under-
63 estimated since animals can acquire impressive orthographic knowledge in the absence of language and
64 phonological skills (Grainger et al., 2012; Rajalingham et al., 2020; Scarf et al., 2016). These findings
65 highlight the urgency to better understand how visual processing and visual attention skills contribute
66 to orthographic learning and self-teaching.

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

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

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

103 Our main contribution in the present study was to implement a more integrated computational model
104 and assess its ability to predict the evolution of eye movements during orthographic learning. For this pur-
105 pose, we started from a recently developed word recognition model, the *BRAID* model (for *Bayesian model*
106 *of word Recognition with Attention, Interference and Dynamics*; Phénix, 2018; Phénix et al., submitted),
107 which includes not only the mechanisms of visual letter similarity and lateral interference classically found
108 in word recognition models, but further the mechanisms of visual acuity and visual attention that are typ-
109 ical of eye movement control models. We extended the *BRAID* model by adding learning mechanisms. As
110 a result, the extended model, called *BRAID-Learn*, features simultaneously the properties of an efficient
111 word recognition model, some of the processes involved in eye movement control, and the mechanisms
112 required for orthographic learning. We then used the *BRAID-Learn* model to predict the evolution of
113 eye movement patterns when being repeatedly exposed to the same set of novel words. A main challenge
114 here was to use the same set of default parameter values that was previously used to simulate a variety of
115 word recognition effects (like frequency and neighborhood effects (Phénix, 2018; Phénix et al., submitted;
116 Phénix et al., 2018), the word superiority or the OVP effect (Phénix, 2018; Phénix et al., submitted; Val-

117 dois et al., submitted), or word length effects (Ginestet et al., 2019; Saghiran et al., 2020)), in an attempt
 118 to account for word recognition, orthographic learning and eye movement data in a single computational
 119 framework. Therefore, overall, our main objective was to explore to what extent a model that was not
 120 specifically designed to account for eye movement control while reading would generalize and predict the
 121 evolution of eye movement patterns during the orthographic learning of novel words.

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

128 **2 The *BRAID-Learn* model**

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

132 2.1 The *BRAID* model

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

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

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

147 2.1.1 The four submodels of the *BRAID* model

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

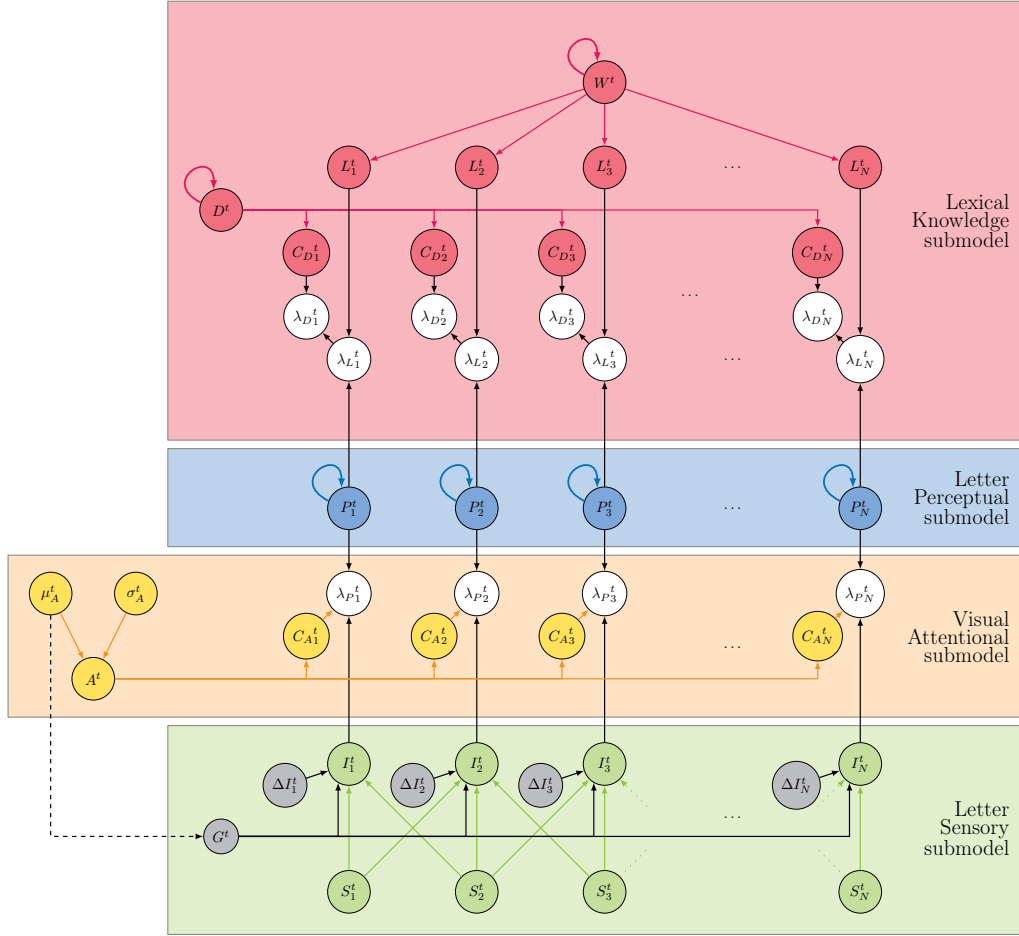


Fig. 1 Graphical representation of the structure of the *BRAID* model. Each of the four colored blocks represents a submodel; each node of the graph represents a variable of the model; and each arrow represents a probability distribution of the model. The graphical schema presented here corresponds to 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.

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

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

161 The “Visual Attentional” submodel Using intermediate variables and probability distributions – techni-
 162 cally, so called “coherence” (Bessi ere et al., 2008; Gilet et al., 2011) and “control” variables (Ph enix,
 163 2018) –, the visual attentional submodel acts as a layer filtering the transfer of bottom-up information,
 164 i.e., from the “Letter Sensory” submodel to the “Letter Perceptual” submodel. This allows to modulate

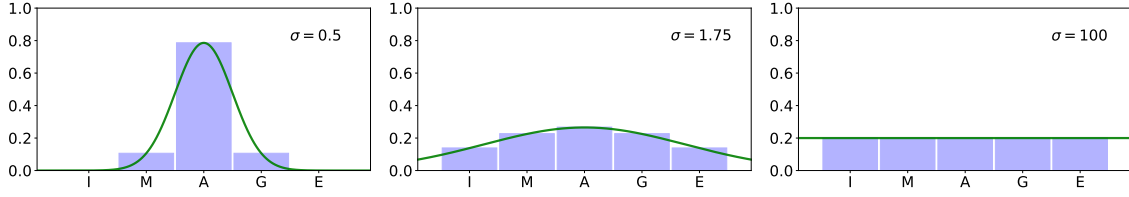


Fig. 2 Illustration of the attention distribution over the letter string of the stimulus word IMAGE for a position of attention $\mu_A = 3$, and for different values of attention dispersion σ_A . The y -axis represents the attention quantity for each position, and the x -axis represents letter positions.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

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

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

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

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

195 Words (w in W) are associated with their corresponding letter sequence $L_{1:N}^{1:T}$ by a probabilistic model,
 196 such that, in each position, the correct letter at that position for this word has a high probability value
 197 (0.974), and all other alternatives have small probability values (0.001).

198 Finally, a third and final Markov chain, over variable D , might be interpreted as a “lexical membership
 199 and word familiarity check”. Variable D is Boolean, with the “True” value representing that a word
 200 stimulus belongs to the known lexicon. The initial, prior distribution $P(D^0)$ is uniform, representing a
 201 50/50 chance that the input stimulus is a known word (a viable assumption to simulate many experimental
 202 setups, although surely not realistic in ecological situations). Variables $D^{1:T}$ are related to Boolean
 203 variables $C_{D_{1:N}}^{1:T}$, in a probabilistic model that represents knowledge about whether a sequence of stimulus
 204 letters corresponds to a known word, or not: for a known word, all variables $C_{D_{1:N}}^{1:T}$ are assumed to be
 205 “True”; on the contrary, for a sequence of stimulus letters that is not a known word, at least one of the
 206 variables $C_{D_{1:N}}^{1:T}$ is assumed to be “False”. These patterns of values serve as templates, to be compared
 207 with values of the coherence variables between the perceptual evidence about letter identity $P_{1:N}^{1:T}$ and
 208 letter sequence $L_{1:N}^{1:T}$, so that “observing” the flow of information between these two variables allows to
 209 infer whether the input stimulus is a known word or not.

210 2.1.2 Probabilistic questions to simulate cognitive tasks

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

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

$$Q_{P_n^T} = P(P_n^T \mid [S_{1:N}^{1:T} = s] [G^{1:T} = g] \mu_A^{1:T} \sigma_A^{1:T} [\lambda_{P_{1:N}}^{1:T} = 1] [\lambda_{L_{1:N}}^{1:T} = 1]) , \quad (1)$$

218 which can be read as: What is the probability distribution over the perceived letter at position n , at
 219 time step T , given the stimulus letter sequence s , gaze position g , the current attentional distribution
 220 (μ_A, σ_A) , and given that information is allowed to propagate from the stimulus to the lexical submodel
 221 $([\lambda_{P_{1:N}}^{1:T} = 1], [\lambda_{L_{1:N}}^{1:T} = 1])$?

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

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

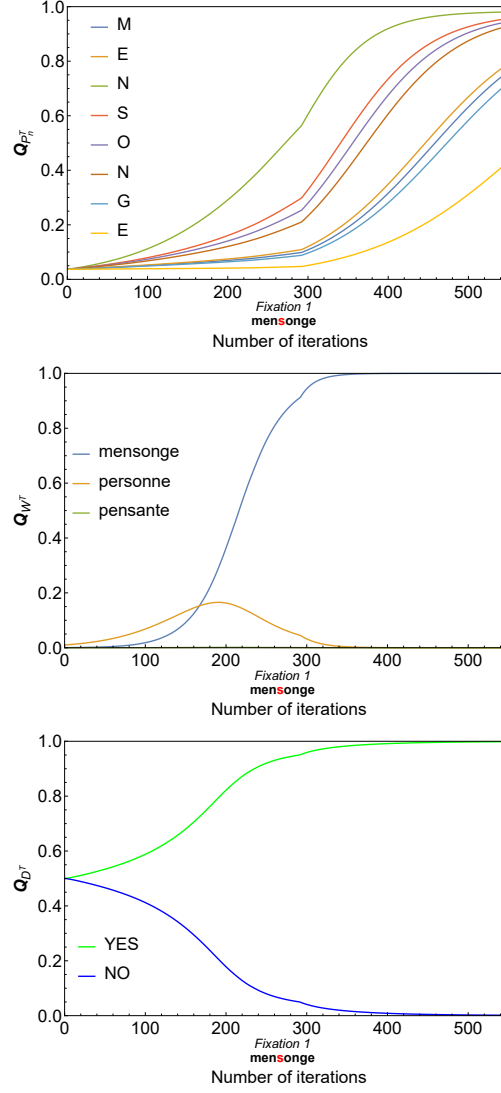


Fig. 3 Evolution of inference for Q_{PT} (top; Eq. (1)), Q_{WT} (middle; Eq. (2)) and Q_{DT} (bottom; Eq. (3)) as a function of simulated time (x -axis) for the 8-letter stimulus *MENSONGE* (*LIE*), with $g = \mu_A = 4$ (eye and attention are positioned over letter “S”, indicated in red under each plot) 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.

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

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

$$Q_{WT} = P(W^T \mid [S_{1:N}^{1:T} = s] [G^{1:T} = g] \mu_A^{1:T} \sigma_A^{1:T} [\lambda_{P_{1:N}}^{1:T} = 1] [\lambda_{L_{1:N}}^{1:T} = 1]) . \quad (2)$$

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

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

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

$$Q_{D^T} = P(D^T \mid [S_{1:N}^{1:T} = s] [G^{1:T} = g] \mu_A^{1:T} \sigma_A^{1:T} [\lambda_{D_{1:N}}^{1:T} = 1] [\lambda_{L_{1:N}}^{1:T} = 1]) . \quad (3)$$

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

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

264 2.2 The *BRAID-Learn* model

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

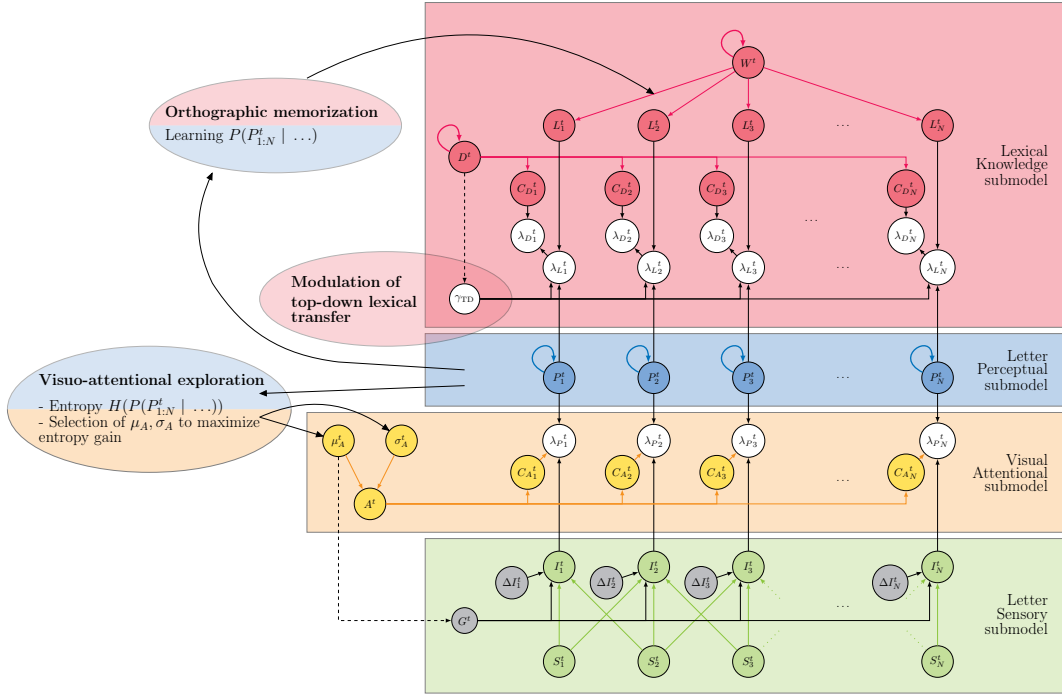


Fig. 4 Graphical representation of the *BRAID-Learn* model. The four colored blocks are the same submodels as in the *BRAID* model, with the same graphical convention (see Fig. 1). To this architecture, the *BRAID-Learn* model adds three mechanisms, represented as colored ovals, that transfer and transform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

274 2.2.1 Efficient accumulation of perceptual evidence about letters

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

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

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

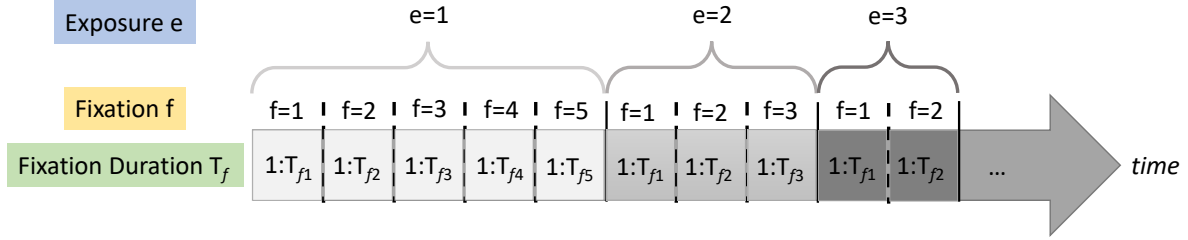


Fig. 5 Schematic representation of the time course of a sequence of several fixations and exposures. Each time a word is encountered (each exposure e), it is fixated once or more times (fixations f), and each such fixation consists in a (variable) duration T_f during which the eye position and the visuo-attentional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

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

$$H(P(X)) = - \sum_X (P(X) \log P(X)) . \quad (4)$$

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

299 The *BRAID-Learn* model aims at optimizing information gain by maximizing entropy decrease. Math-
 300 ematically, before fixation $f + 1$, we enumerate a range of possible values for upcoming attention position
 301 $\mu_A^{T,f+1,e}$ and dispersion $\sigma_A^{T,f+1,e}$; for each such possible future attention distribution, and assuming that
 302 the input stimulus will not change during next fixation, we simulate letter recognition in each position n
 303 with

$$P_{\text{next}}(n, \mu_A^{f+1,e}, \sigma_A^{f+1,e}) = P(P_n^{T,f+1,e} | [S_n^{T,f+1,e} = s] [G^{T,f+1,e} = g^{f+1,e}] \mu_A^{T,f+1,e} \sigma_A^{T,f+1,e}) . \quad (5)$$

304 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
 305 can then compute the entropy gain between the predicted and current distribution over letters, for all pos-
 306 sible attention distribution parameters and average it across positions; we note this $\overline{\Delta H}(\mu_A^{T,f+1,e}, \sigma_A^{T,f+1,e})$.

307 To model the physical “motor cost” of performing the visuo-attentional displacement to each enumer-
 308 ated future fixations, we use a straightforward measure, considering only the magnitude of the supposed
 309 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
 310 large displacements of gaze and attentional positions, so that the overall gain measure TG that the model
 311 maximizes is a weighted combination of information gain penalized by motor cost:

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

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

314 Having described how, at any point in time, the next fixation parameters are selected, we define the
 315 initial parameters and termination criterion. Whatever the stimulus, whether it is a word or not, and

316 since the model, at initialization, has no knowledge of the stimulus type, the parameters for the first
 317 fixation are identical.

318 Therefore, in the context of current experiments, that only deal with 8-letter long stimuli, we assume
 319 that gaze and attention “land” at position $\mu_A^{T,1,e} = 3$ whatever the exposure e . This initial position is the
 320 rounded value closest to the one from our previous experimental observations (3.01 across all item types,
 321 i.e., for words and pseudowords, and across all repetition exposures) in which expert readers had to read
 322 8-letter words and pseudowords (Ginestet et al., 2020). For the initial dispersion of visual attention, we
 323 apply the usual default value in the *BRAID* model: $\sigma_A^{T,1,e} = 1.75$.

324 We define two termination criteria: the first defines how long each fixation is going to last, and the
 325 second is used to decide that no further fixations are going to be performed. Concerning fixation duration,
 326 we assume that the model aims at having as short fixations as possible (later on, during data analyses,
 327 back-to-back fixations on the same spatial position are aggregated and counted as a single fixation on this
 328 position; aiming for short fixations is not a theoretical claim, instead it just yields temporal granularity in
 329 our simulations). Initial simulations have shown that, in the first few iterations of the predictive evaluation
 330 of entropy gain, the “winning parameters” were numerically close, until a clear set of value emerged and,
 331 most of the time, stayed ahead until the maximal window of predictive computation. This maximal time
 332 is set, for current experiments, at $T = 290$ iterations, well above the average fixation duration reported for
 333 novel words in behavioral experiments (Ginestet et al., 2020; Joseph et al., 2014; Pellicer-Sanchez, 2016).
 334 We thus detect the time-step T_f for which the predicted winning parameter values have been stable for
 335 20 previous time-steps. Finally, we set the minimal duration T_f to be at 50 iterations. The upcoming
 336 fixation is then performed with these winning parameters for that duration.

337 The second termination criterion prevents a further fixation when its expected information gain is
 338 below a threshold. Since fixations are of varying duration T_f , this is scaled as a function of T_f . We have
 339 empirically calibrated our stop criterion to correspond to 1 nat of information gain for the whole word,
 340 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
 341 (our simulations use natural base e for entropy calculation, which is therefore measured in nats instead of
 342 bits). Therefore, our termination threshold is $T_f/(N \times 250)$; whenever a fixation is selected and associated
 343 to an information gain below this threshold, it is not performed by the model, and the current exposure
 344 e is considered terminated.

345 2.2.2 Modulation of lexical influence during word learning

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

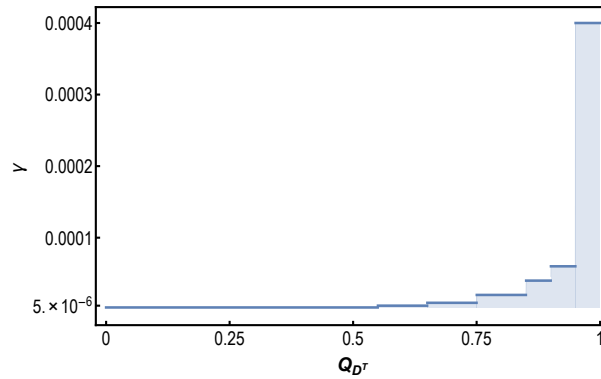


Fig. 6 Graph of the function of parameter γ (y -axis), that pilots the amount of top-down transfer of lexical information, as a function of Q_{D^T} , the probability that $D^T = \text{true}$ (x -axis), computed by Eq. (3), that represents probability of lexical membership and word familiarity.

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

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

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

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

385 2.2.3 Memorization and update of orthographic traces

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

392 In the first case, that is, for updating a lexical distribution, at the end of exposure e , and for each posi-
 393 tion n , the complete probability distribution about the perceived letter, $P(P_n^{T,f,e} | [S_n^{T,f,e} = s] [G^{T,f,e} =$
 394 $g] \mu_A \sigma_A)$, is combined with the previous probability distribution about the letter at that position,
 395 $P(L_n^e | [W^e = w])$, in the lexical sub-model, for the recognized word w :

$$P(L_n^{e+1} | [W^{e+1} = w]) = \left[P(L_n^e | [W^e = w]) \cdot (e - 1) + P(P_n^{T,f,e} | [S_n^{T,f,e} = s] [G^{T,f,e} = g] \mu_A \sigma_A) \right] / e \quad (7)$$

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

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

402 2.3 Summary

403 The *BRAID-Learn* model includes three mechanisms that affect letter identification within strings during
 404 word recognition, namely an acuity gradient, a mechanism of lateral interference between adjacent letters
 405 and a visual attention filter. The model assumes that a novel word trace is created each time the input
 406 letter-string is detected as not belonging to the model lexical knowledge. Furthermore, detecting that the
 407 input is novel entails decreasing the top-down feedback from word knowledge to letter perception; this
 408 yields a relative increase in the effect of perceptual evidence about letters from bottom-up processing. In
 409 other words, bottom-up information is privileged as the principal source of information on letter identity.
 410 Visuo-attentional exploration during processing is defined by a mathematical principle of entropy gain
 411 maximisation. The entropy gain maximisation principle allows selecting the visuo-attentional distribution

412 parameters – attentional focus and dispersion – more likely to speed-up accumulation of perceptual infor-
413 mation about letters. This mechanism leads the model to realize as many visuo-attentional displacements
414 as necessary as long as perceptual information is not precise enough. Visuo-attentional exploration is
415 further constrained by a motor-cost parameter that penalizes large displacements over the letter-string.
416 When visuo-attentional exploration is terminated, lexical knowledge is updated. This final mechanism
417 simulates either the reinforcement of the orthographic representation of a known word or the creation of
418 a new lexical trace, both reflecting orthographic learning.

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

424 **3 Simulation of orthographic learning: the effect of repeated reading of novel words on eye** 425 **movements**

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

439 **3.1 Simulating orthographic learning: an illustrative example**

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

445 3.1.1 Applying *BRAID-Learn* to a known word

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

450 The simulation yielded two fixations for processing the word *MENSONGE*. The first one was dictated
 451 by default parameters of the *BRAID-Learn* model: whatever the word type, the first fixation for an 8-
 452 letter long stimulus is at position $g = \mu_A = 3$ (over the “N” of *MENSONGE*), with attentional dispersion
 453 $\sigma_A = 1.75$, and lasts 290 iterations. The second fixation, selected by optimizing the predicted perceptual
 454 information gain, was at position 7 (over the “G” of *MENSONGE*), with attentional dispersion $\sigma_A = 2.0$,
 455 and lasts 250 iterations.

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

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

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

475 3.1.2 Applying *BRAID-Learn* to a novel word

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

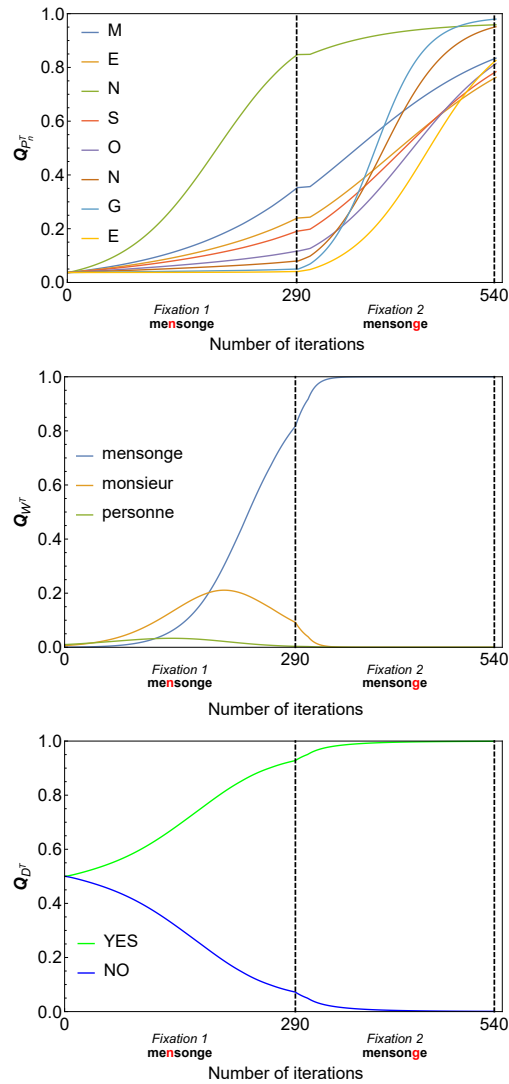


Fig. 7 Evolution of inference for Q_{PT} (top; Eq. (1)), Q_{WT} (middle; Eq. (2)) and Q_{DT} (bottom; Eq. (3)) as a function of simulated time (x -axis) for the 8-letter stimulus *MENSONGE*, with fixations computed by the *BRAID-Learn* model. Graphical representation is identical to the one of Fig. 3,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

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

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

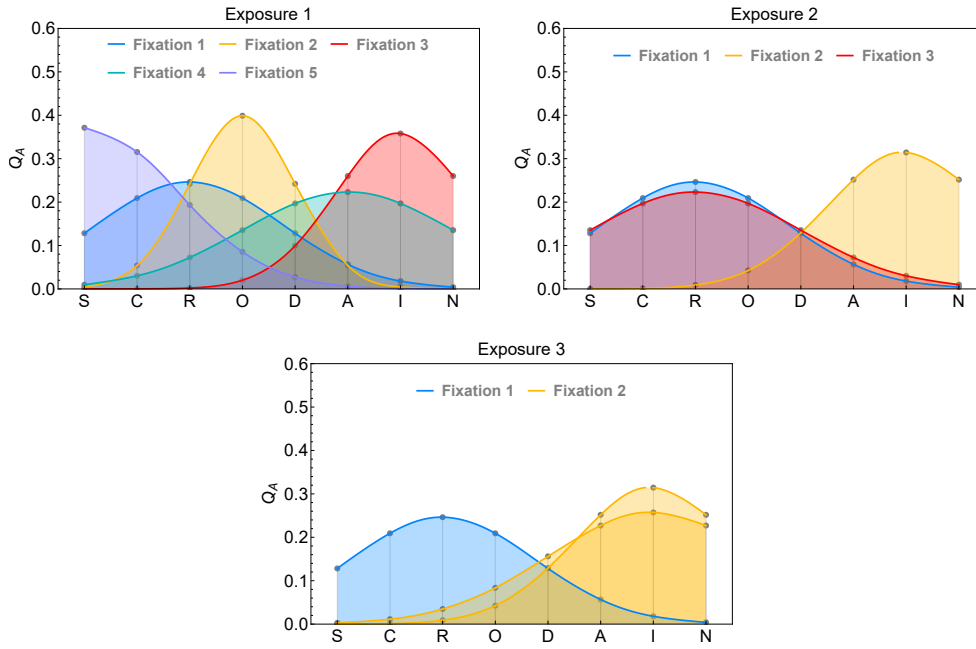


Fig. 8 Evolution of visuo-attentional parameters selected by the *BRAID-Learn* model during the first three exposures to the 8-letter stimulus non-word *SCRODAIN*. Each plot represents the probability value Q_A attributed to each position by the attentional model, following a Gaussian probability distribution. 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, $\sigma_A^{T,f+1,e}$.

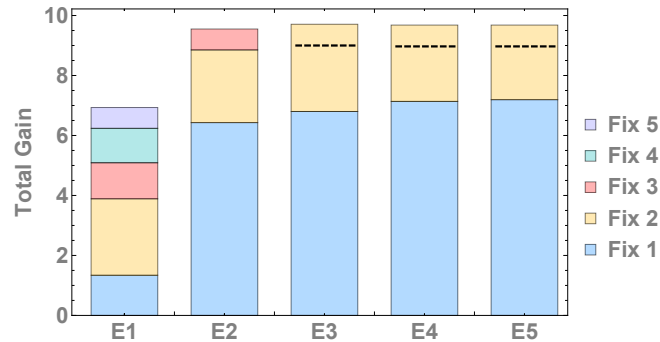


Fig. 9 Total Gain as a function of exposure number (noted E1 to E5) and fixation number (noted Fix 1 to Fix 5), during the processing of the 8-letter stimulus non-word *SCRODAIN*. The dashed horizontal line represents subsequent fixations that occur on the same spatial position, and are thus aggregated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

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

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

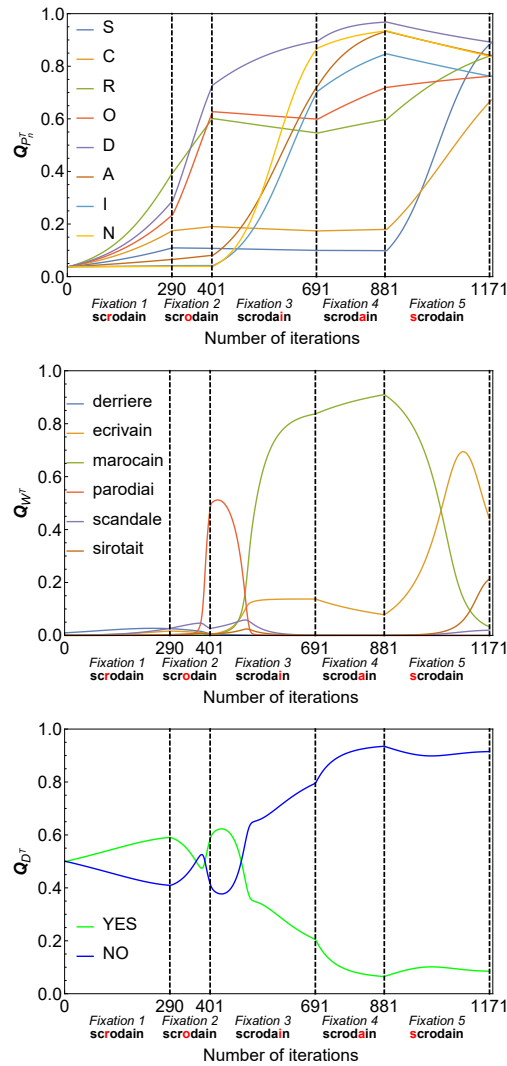


Fig. 10 Evolution of inference for $Q_{P_n^T}$ (top; Eq. (1)), Q_{W^T} (middle; Eq. (2)) and Q_{D^T} (bottom; Eq. (3)) as a function of simulated time (x -axis) for the first exposure to the 8-letter stimulus non-word *SCRODAIN*, with fixations computed by the *BRAID-Learn* model. Graphical representation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

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

501 Since the stimulus was recognized as a new word, the probability distribution over words was switching
502 between hypotheses, with no clear convergence to a single, winning hypothesis. This is the expected
503 behavior, since, during first exposure to a novel word, the word space W does not contain a point
504 corresponding to the stimulus. Instead, the most likely hypotheses in word space were close competitors
505 to the stimulus, with the best one depending on processing stage, and more specifically, depending on
506 current letter perception and fixation position. For example, consider iteration 401: few letters were well
507 identified and gaze and attention were centered on the 4th position (the “O” of *SCRODAIN*). At this
508 point, the most probable word was *PARODIAI* (*PARODIED*), which shares with *SCRODAIN* the “R”,
509 “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.

547 3.2.2 Simulated results

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

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

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

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

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

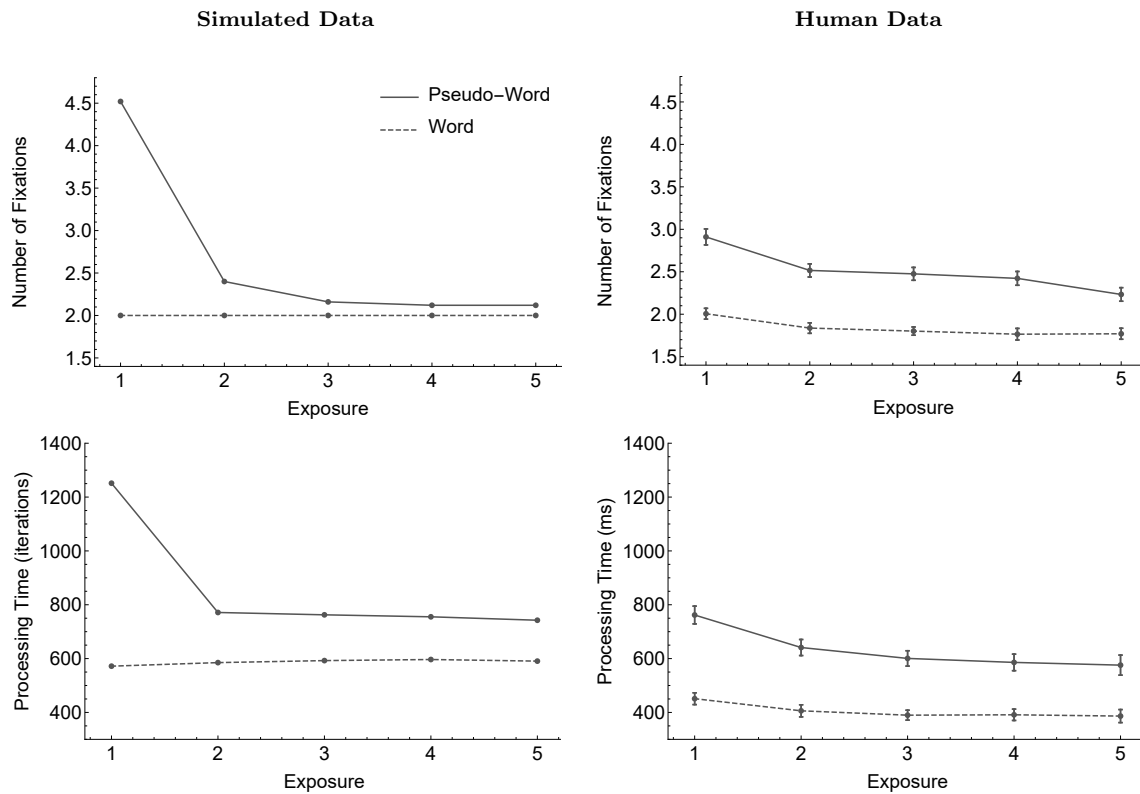


Fig. 11 Number of fixations (top plots) and processing time (bottom plots), as a function of exposures, from the reading aloud task in the experiment of Ginestet et al. (2020) (right plots) and simulated by the *BRAID-Learn* model (left plots), for known word stimuli (dashed lines) and novel word stimuli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

584 In all models, contrasts were specified as 0.5/-0.5 or 2/1/0/-1/-2 when independent variables have,
 585 respectively, 2 or 5 modalities. All statistical models and simulated results are provided in Supplementary
 586 Material¹ (for a quick access, see the .html file from “Statistical_files” folder).

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

592 We first report statistical analyses concerning the number of fixations. Results showed main effects
 593 of item type ($\beta = -0.66$, $z = -3.57$, $p < .001$) and number of exposures ($\beta = 0.22$, $z = 3.37$, $p <$
 594 $.001$) on the number of fixations. As in the experimental data, the number of fixations across exposures
 595 decreased faster for novel words than for known words ($\beta = -0.44$, $z = -3.37$, $p < .001$). Post-
 596 hoc comparisons showed that this decrease mainly occurred between the first and the second exposure
 597 ($\beta = 2.12$, $z = 3.31$, $p < .001$; non significant interaction between the second and third exposure:
 598 $\beta = 0.24$, $z = 0.43$, $p = .669$).

599 We now turn to analyses of processing times. Results showed main effects of item type ($\beta = -265.98$, $t =$
 600 -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

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

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

612 4 Discussion

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

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

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

626 4.1 Theoretical contribution of the *BRAID-Learn* model

627 Our main contribution in the present paper is to make new assumptions about the mechanisms involved
628 in the acquisition of orthographic knowledge and describe *BRAID-Learn*, the first computational model
629 where the central focus is orthographic learning. Although the models of reading acquisition developed
630 within the self-teaching framework (Pritchard et al., 2018; Ziegler et al., 2014) were designed to be able
631 to enrich their orthographic knowledge through the acquisition of new orthographic representations, they
632 did not implement the visuo-orthographic mechanisms involved in novel word orthographic learning.
633 These models were derived from models of reading aloud, thus placing emphasis on phonological skills. In
634 contrast, *BRAID-Learn* is the extension of a word recognition model. As a result, *BRAID-Learn* is explicit
635 on the mechanisms of visual and visual attentional processing that are involved in the identification of
636 the input word letter-string. However, the model does not include any of the phonological components

637 usually postulated to account for reading aloud and reading acquisition. Thus, according to *BRAID-Learn*,
638 orthographic learning is mainly conditional on the efficiency with which letters are identified within the
639 novel word input string, while it is mainly conditional on successful phonological decoding according to
640 self-teaching models.

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

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

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

672 Last, the online modulation of top-down lexical influence allowed configuring the model so that,
673 from the same mathematical principle of perceptual evidence gain maximization, it would yield different
674 oculomotor behaviors for known words and novel (or pseudo-) words. Evidence that different reading
675 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

789 attention captures, and thus refixations. This translates in a larger number of fixations and longer total
790 processing time for novel words than for known words, since only the latter benefit from top-down lexical
791 influence that boosts letter identification from the first encounter. However, top-down lexical information
792 becomes effective from the second exposure to the same novel word. Information extracted on letter
793 identity and memorized during the first exposure can then be used to speed-up letter processing. This
794 positive effect of top-down influence from the newly acquired word orthographic representation results in
795 large decreases of both number of fixations and processing time at the second exposure. Further exposures
796 lead to further improvements of the orthographic representation of the word being acquired and variations
797 in the strength of lexical feedback result in more gradual changes in eye movements. Overall, the model
798 behavior is well in line with the empirical findings reported in humans, namely that the first exposure
799 to a novel word is more critical than later ones for orthographic learning and that the orthographic
800 representations of new words are stabilized, after only a small, single-digit number of exposures (Ginestet
801 et al., 2020; Nation et al., 2007; Pellicer-Sanchez, 2016; Share, 2004).

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

820 The simulations were carried out to assess to what extent a model limited to visual orthographic pro-
821 cessing would account for the features of eye movement patterns reported in humans during orthographic
822 learning. The good qualitative account of eye movement pattern changes across exposures suggests that
823 the two mechanisms of visual attention and lexical feedback postulated by the model are critical factors
824 for orthographic learning in humans, and may be also in animals (Grainger et al., 2012; Scarf et al.,
825 2016). It is noteworthy that *BRAID-Learn* was not specifically developed to account for eye movement
826 patterns while reading. As such, *BRAID-Learn*, for the first time, offers a unified account of three aspects
827 of the reading system that are typically modelled independently, namely word recognition, eye movement

828 control and orthographic learning. Visual attention appears critical for an integrated account of these
829 different dimensions of the reading process.

830 **Acknowledgments**

831 This work has been supported by a French Ministry of Research (MESR) grant to EG and the French gov-
832 ernment as part of the e-FRAN “FLUENCE” project funded by the “Investissement d’Avenir” program
833 handled by the “Caisse des Dépôts et Consignations”.

834 **Conflict of Interest Statement**

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

837 **Open Practices Statement**

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

840

841 **Appendix A: List of items**

842 Lists of Items used in simulations

843 Pseudo-words : broufand, chaiquau, deinrint, faingion, gouciant, nauplois, ploirtart, speirain, quinsard,
844 tramoint, ceitteau, chanquet, coirtint, drottont, flommais, glounein, priquoïn, quarlant, siampoie, trim-
845 pond, bussiond, cherrein, ciercard, claffand, fentroit, phacrait, prinnant, tauppart, scrodain, trancare.

846 Control words : uniforme, portrait, enceinte, mouchoir, surprise, complice, scandale, chanteur, immeu-
847 ble, avantage, physique, revanche, horrible, boutique, sensible, fauteuil, chocolat, mensonge, solution,
848 voyageur, prochain, grandeur, nocturne, lointain, religion, empereur, division, quartier, province, juge-
849 ment.

850 **Appendix B: Item-level simulations**

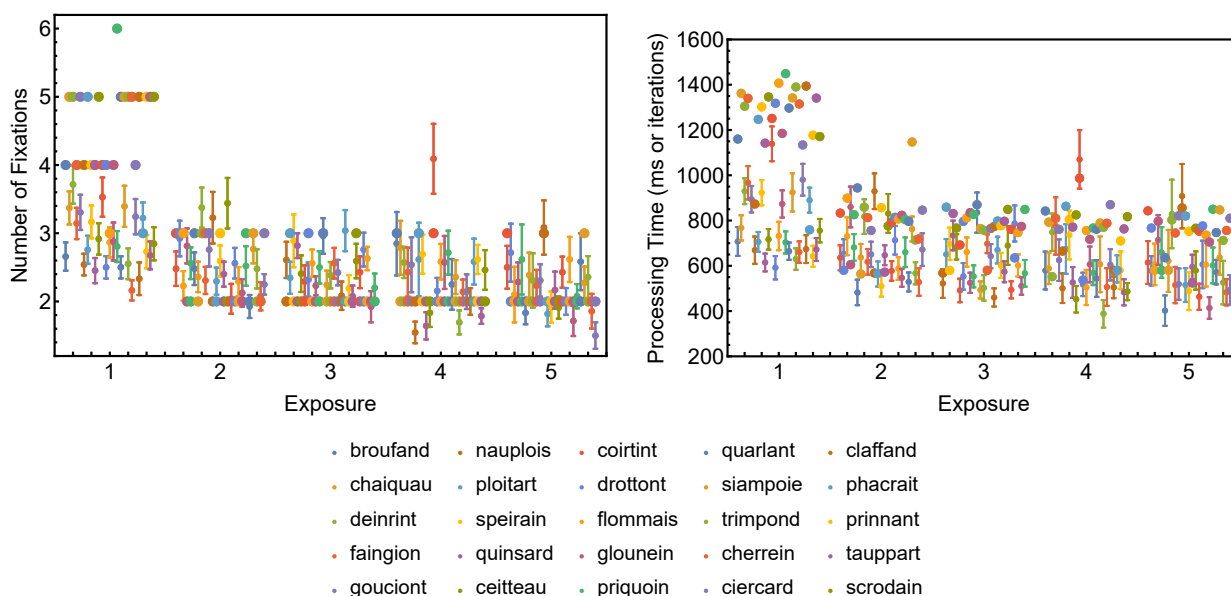


Fig. A.1 Number of Fixations (left plot) and Processing Time (right plot) reported as a function of exposures and for each novel word learned by the *BRAID-Learn* model. Smaller dots and their vertical bars (standard errors) represent behavioral data; larger dots represent simulated results.

851 References

- 852 Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic
 853 word reading. *Psychological Review*, *105*(4), 678–723. [https://doi.org/10.1037/0033-295X.105.4.](https://doi.org/10.1037/0033-295X.105.4.678-723)
 854 678-723
- 855 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory
 856 hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- 858 Bernard, J.-B., del Prado Martin, F. M., Montagnini, A., & Castet, E. (2008). A model of optimal
 859 oculomotor strategies in reading for normal and damaged visual fields. *Deuxième conférence*
 860 *française de Neurosciences Computationnelles*, “*Neurocomp08*”.
- 861 Besner, D., Risko, E. F., Stolz, J. A., White, D., Reynolds, M., O’Malley, S., & Robidoux, S. (2016).
 862 Varieties of attention: Their roles in visual word identification. *Current Directions in Psychological*
 863 *Science*, *25*(3), 162–168. <https://doi.org/10.1177/0963721416639351>
- 864 Bessièrè, P., Laugier, C., & Siegwart, R. (Eds.). (2008). *Probabilistic reasoning and decision making in*
 865 *sensory-motor systems* (Vol. 46). Springer. <https://doi.org/10.1007/978-3-540-79007-5>
- 866 Bosse, M.-L. (2005). De la relation entre acquisition de l’orthographe lexicale et traitement visuo-
 867 attentionnel chez l’enfant. *Rééducation orthophonique*, *222*, 9–30.
- 868 Bosse, M.-L., Chaves, N., Largy, P., & Valdois, S. (2015). Orthographic learning during reading: The
 869 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>
 870

- 871 Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span
872 deficit hypothesis. *Cognition*, *104*(2), 198–230. <https://doi.org/10.1016/j.cognition.2006.05.009>
- 873 Bosse, M.-L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: A
874 cross-sectional study. *Journal of Research in Reading*, *32*(2), 230–253. [https://doi.org/10.1111/
875 j.1467-9817.2008.01387.x](https://doi.org/10.1111/j.1467-9817.2008.01387.x)
- 876 Bowey, J., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders'
877 silent reading: A critical test of the self-teaching hypothesis. *Journal of Experimental Child Psy-
878 chology*, *92*(2), 203–219. <https://doi.org/10.1016/j.jecp.2005.06.005>
- 879 Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From
880 inkmarks to ideas: Current issues in lexical processing* (pp. 151–179). Psychology Press.
- 881 Castles, A., & Nation, K. (2008). Learning to be a good orthographic reader. *Journal of Research in
882 Reading*, *31*(1), 1–7. <https://doi.org/10.1111/j.1467-9817.2007.00367.x>
- 883 Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice
884 to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. [https://doi.org/10.1177/
885 1529100618772271](https://doi.org/10.1177/1529100618772271)
- 886 Chaffin, R., Morris, R., & Seely, R. E. (2001). Learning new word meanings from context: A study of
887 eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1),
888 225–235. <https://doi.org/10.1037/0278-7393.27.1.225>
- 889 Chan, K. S.-C., & Yeung, P.-S. (2020). Prediction of chinese reading fluency by verbal and non-verbal
890 visual attention span measures. *Frontiers in Psychology*, (10), 3049. [https://doi.org/10.3389/
891 fpsyg.2019.03049](https://doi.org/10.3389/fpsyg.2019.03049)
- 892 Chaves, N., Totereau, C., & Bosse, M.-L. (2012). Acquérir l'orthographe lexicale: Quand savoir lire ne
893 suffit pas. *ANAE – Approche Neuropsychologique des Apprentissages chez L'Enfant*, *118*, 271–
894 279.
- 895 Chen, N. T., Zheng, M., & Ho, C. S.-H. (2019). Examining the visual attention span hypothesis in chinese
896 developmental dyslexia. *Reading and Writing*, *32*, 639–662. [https://doi.org/10.1007/s11145-018-
897 9882-1](https://doi.org/10.1007/s11145-018-9882-1)
- 898 Colas, F., Flacher, F., Tanner, T., Bessi re, P., & Girard, B. (2009). Bayesian models of eye movement
899 selection with retinotopic maps. *Biological Cybernetics*, *100*(3), 203–214. [https://doi.org/10.
900 1007/s00422-009-0292-y](https://doi.org/10.1007/s00422-009-0292-y)
- 901 Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded
902 model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.
903 <https://doi.org/10.1037/0033-295X.108.1.204>
- 904 Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children
905 self-teach? *Journal of Experimental Child Psychology*, *95*, 56–77. [https://doi.org/10.1016/j.jecp.
906 2006.03.008](https://doi.org/10.1016/j.jecp.2006.03.008)
- 907 Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of
908 orthographic processing. *Reading and Writing: An interdisciplinary Journal*, *14*, 549–568. <https://doi.org/10.1023/A:1011100226798>
909

- 910 Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*(3),
911 713. <https://doi.org/10.1037/a0019738>
- 912 de Jong, P. F., Bitter, D. J., van Setten, M., & Marinus, E. (2009). Does phonological recoding occur
913 during silent reading and is it necessary for orthographic learning? *Journal of Experimental Child*
914 *Psychology*, *104*(3), 267–282. <https://doi.org/10.1016/j.jecp.2009.06.002>
- 915 Dubois, M., Kyllingsbæk, S., Prado, C., Musca, S. C., Peiffer, E., Lassus-Sangosse, D., & Valdois, S.
916 (2010). Fractionating the multi-character processing deficit in developmental dyslexia: Evidence
917 from two case studies. *Cortex*, *46*, 717–738. <https://doi.org/10.1016/j.cortex.2009.11.002>
- 918 Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based
919 on spatially distributed lexical processing. *Vision research*, *42*(5), 621–636. [https://doi.org/10.1016/S0042-6989\(01\)00301-7](https://doi.org/10.1016/S0042-6989(01)00301-7)
- 920 Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). Swift: A dynamical model of saccade
921 generation during reading. *Psychological review*, *112*(4), 777. <https://doi.org/10.1037/0033-295X.112.4.777>
- 922 Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C.
923 (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840
924 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. <https://doi.org/10.3758/BRM.42.2.488>
- 925 Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as
926 experiments. *Frontiers in psychology*, *3*, 151. <https://doi.org/10.3389/fpsyg.2012.00151>
- 927 Gerbier, E., Bailly, G., & Bosse, M.-L. (2015). Using Karaoke to enhance reading while listening: Impact on
928 word memorization and eye movements. *Speech and Language Technology for Education (SLaTE)*,
929 59–64.
- 930 Gerbier, E., Bailly, G., & Bosse, M.-L. (2018). Audio-visual synchronization in reading while listening to
931 texts: Effects on visual behavior and verbal learning. *Computer Speech and Language*, *47*, 79–92.
932 <https://doi.org/10.1016/j.csl.2017.07.003>
- 933 Geyer, L. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics*,
934 *22*(5), 487–490. <https://doi.org/10.3758/BF03199515>
- 935 Gilet, E., Diard, J., & Bessière, P. (2011). Bayesian action-perception computational model: Interaction
936 of production and recognition of cursive letters. *PLoS ONE*, *6*(6), e20387. <https://doi.org/10.1371/journal.pone.0020387>
- 937 Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical
938 decision: The role of visual attention. *Vision Research*, *159*, 10–20. <https://doi.org/10.1016/j.visres.2019.03.003>
- 939 Ginestet, E., Valdois, S., Diard, J., & Bosse, M.-L. (2020). Orthographic learning of novel words in adults:
940 Effects of exposure and visual attention on eye movements. *Journal of Cognitive Psychology*,
941 *32*(8), 785–804. <https://doi.org/10.1080/20445911.2020.1823987>
- 942 Gomez, P., Ratcliff, R., & Perea, M. (2008). The Overlap model: A model of letter position coding.
943 *Psychological Review*, *115*(3), 577–601. <https://doi.org/10.1037/a0012667>
- 944

- 949 Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in
950 Baboons (*Papio Papio*). *Science*, *336*(6078), 245–248. <https://doi.org/10.1126/science.1218152>
- 951 Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception:
952 Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human*
953 *Perception and Performance*, *20*(6), 1158. <https://doi.org/10.1037/0096-1523.20.6.1158>
- 954 Howard, D., & Best, W. (1996). Developmental phonological dyslexia: Real word reading can be com-
955 pletely normal. *Cognitive Neuropsychology*, *13*(6), 887–934. <https://doi.org/10.1080/026432996381854>
- 956 Inserm. (2007). Dyslexie, dysorthographe, dyscalculie. Bilan des données scientifiques. *Paris: Les éditions*
957 *Inserm*, *15*, 842p–Expertise collective.
- 958 Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role
959 of context. *Journal of Experimental Child Psychology*, *166*, 190–211. <https://doi.org/10.1016/j.jecp.2017.08.010>
- 960 Joseph, H., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements
961 reveal order of acquisition effects following incidental exposure to new words during silent reading.
962 *Cognition*, *133*(1), 238–248. <https://doi.org/10.1016/j.cognition.2014.06.015>
- 963 Kyte, C. S., & Johnson, C. J. (2009). The role of phonological recoding in orthographic learning. *Journal*
964 *of Experimental Child Psychology*, *93*(2), 166–185. <https://doi.org/10.1016/j.jecp.2005.09.003>
- 965 Laberge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading.
966 *Cognitive Psychology*, *6*(10), 293–323. [https://doi.org/10.1016/0010-0285\(74\)90015-2](https://doi.org/10.1016/0010-0285(74)90015-2)
- 967 Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): Still no identifi-
968 cation without attention. *Psychological Review*, *111*(4), 880–913. [https://doi.org/10.1037/0033-](https://doi.org/10.1037/0033-295X.111.4.880)
969 [295X.111.4.880](https://doi.org/10.1037/0033-295X.111.4.880)
- 970 Lee, T. S., & Stella, X. Y. (2000). An information-theoretic framework for understanding saccadic eye
971 movements. *Advances in neural information processing systems*, 834–840.
- 972 Legge, G. E., Hooven, T. A., Klitz, T. S., Mansfield, J. S., & Tjan, B. S. (2002). Mr. Chips 2002: New
973 insights from an ideal-observer model of reading. *Vision research*, *42*(18), 2219–2234. [https://doi.org/10.1016/S0042-6989\(02\)00131-1](https://doi.org/10.1016/S0042-6989(02)00131-1)
- 974 Legge, G. E., Klitz, T. S., & Tjan, B. S. (1997). Mr. Chips: An ideal-observer model of reading. *Psycho-*
975 *logical review*, *104*(3), 524. <https://doi.org/10.1037/0033-295X.104.3.524>
- 976 Lobier, M., Dubois, M., & Valdois, S. (2013). The role of visual processing speed in reading speed
977 development. *PLoS ONE*, *8*(4), e58097. <https://doi.org/10.1371/journal.pone.0058097>
- 978 Lowell, R., & Morris, R. K. (2014). Word length effects on novel words: Evidence from eye movements.
979 *Attention, Perception, & Psychophysics*, *76*(1), 179–189. [https://doi.org/10.3758/s13414-013-](https://doi.org/10.3758/s13414-013-0556-4)
980 [0556-4](https://doi.org/10.3758/s13414-013-0556-4)
- 981 Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. D. (2018). Interaction in spoken word
982 recognition models: Feedback helps. *Frontiers in Psychology*, *9*, 369. [https://doi.org/10.3389/](https://doi.org/10.3389/fpsyg.2018.00369)
983 [fpsyg.2018.00369](https://doi.org/10.3389/fpsyg.2018.00369)

- 986 Marinelli, C. V., Zoccolotti, P., & Romani, C. (2020). The ability to learn new written words is modulated
987 by language orthographic consistency. *PloS one*, *15*(2), e0228129. [https://doi.org/10.1371/
988 journal.pone.0228129](https://doi.org/10.1371/journal.pone.0228129)
- 989 McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in
990 letter perception: Part 1. an account of basic findings. *Psychological Review*, *88*(5), 375–407.
991 <https://doi.org/10.1037/0033-295X.88.5.375>
- 992 Mozer, M. C., & Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge:
993 A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, *2*(2), 96–123.
994 <https://doi.org/10.1162/jocn.1990.2.2.96>
- 995 Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*,
996 *434*(7031), 387. <https://doi.org/10.1038/nature03390>
- 997 Nation, K., Angell, P., & Castles, A. (2007). Orthographic learning via self-teaching in children learning
998 to read English: Effects of exposure, durability, and context. *Journal of Experimental Child
999 Psychology*, *96*(1), 71–84. <https://doi.org/10.1016/j.jecp.2006.06.004>
- 1000 Nation, K., & Castles, A. (2017). Putting the learning into orthographic learning. In K. Cain, D. L.
1001 Compton, & R. K. Parrila (Eds.), *Theories of reading development* (pp. 147–168). John Benjamins
1002 Publishing Company.
- 1003 Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex
1004 perceptual environments. *Proceedings of the National Academy of Sciences*, *107*(11), 5232–5237.
1005 <https://doi.org/10.1073/pnas.0911972107>
- 1006 Niolaki, G. Z., Vousden, J., Terzopoulos, A. R., Taylor, L. M., Hall, S., & Masterson, J. (2020). Predictors
1007 of single word spelling in english speaking children: A cross-sectional study. *Journal of Research
1008 in Reading*, *43*(4), 577–596. <https://doi.org/10.1111/1467-9817.12330>
- 1009 Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal bayesian decision
1010 process. *Psychological Review*, *113*(2), 327–357. <https://doi.org/10.1037/0033-295X.113.2.327>
- 1011 Norris, D., McQueen, J., & Cutler, A. (2018). Commentary on “interaction in spoken word recognition
1012 models”. *Frontiers in Psychology*, *9*, 1568. <https://doi.org/10.3389/fpsyg.2018.01568>
- 1013 Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification
1014 model for letter and word recognition: The word-superiority effect. *Psychological Review*, *89*(5),
1015 573–594. <https://doi.org/10.1037/0033-295X.89.5.573>
- 1016 Pagan, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing and retrieval
1017 effects in adults. *Cognitive Science*, *43*. <https://doi.org/10.1111/cogs.12705>
- 1018 Pellicer-Sanchez, A. (2016). Incidental l2 vocabulary acquisition from and while reading. *Studies in Second
1019 Language Acquisition*, *38*, 97–130. <https://doi.org/10.1017/S0272263115000224>
- 1020 Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of com-
1021 putational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315.
1022 <https://doi.org/10.1037/0033-295X.114.2.273>

- 1023 Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading
1024 aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*(2), 106–
1025 51. <https://doi.org/10.1016/j.cogpsych.2010.04.001>
- 1026 Perry, C., Zorzi, M., & Ziegler, J. C. (2019). Understanding dyslexia through personalized large-scale com-
1027 putational models. *Psychological science*, *30*(3), 386–395. <https://doi.org/10.1177/0956797618823540>
- 1028 Phénix, T. (2018). *Modélisation bayésienne algorithmique de la reconnaissance visuelle de mots et de*
1029 *l'attention visuelle* (Doctoral dissertation). Univ. Grenoble Alpes.
- 1030 Phénix, T., Ginestet, E., Valdois, S., & Diard, J. (submitted). Visual attention matters during word
1031 recognition: A Bayesian modeling approach. —.
- 1032 Phénix, T., Valdois, S., & Diard, J. (2018). Reconciling opposite neighborhood frequency effects in lexical
1033 decision: Evidence from a novel probabilistic model of visual word recognition. In T. Rogers, M.
1034 Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual conference of the cognitive*
1035 *science society* (pp. 2238–2243). Cognitive Science Society.
- 1036 Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A computational model of the self-
1037 teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive Science*, 1–49.
1038 <https://doi.org/10.1111/cogs.12571>
- 1039 R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Sta-
1040 tistical Computing. Vienna, Austria. <https://www.R-project.org/>
- 1041 Raj, R., Geisler, W. S., Frazor, R. A., & Bovik, A. C. (2005). Contrast statistics for foveated visual
1042 systems: Fixation selection by minimizing contrast entropy. *JOSA A*, *22*(10), 2039–2049. <https://doi.org/10.1364/JOSAA.22.002039>
- 1044 Rajalingham, R., Kar, K., Sanghavi, S., Dehaene, S., & DiCarlo, J. J. (2020). The inferior temporal
1045 cortex is a potential cortical precursor of orthographic processing in untrained monkeys. *Nature*
1046 *Communications*, *11*(1), 1–13. <https://doi.org/10.1038/s41467-020-17714-3>
- 1047 Rau, A. K., Moll, K., Snowling, M. J., & Landerl, K. (2015). Effects of orthographic consistency on eye
1048 movement behavior: German and English children and adults process the same words differently.
1049 *Journal of Experimental Child Psychology*, *130*, 92–105. [https://doi.org/10.1016/j.jecp.2014.09.](https://doi.org/10.1016/j.jecp.2014.09.012)
1050 012
- 1051 Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for
1052 initial fixation locations and refixations within the E-Z Reader model. *Vision research*, *39*(26),
1053 4403–4411. [https://doi.org/10.1016/S0042-6989\(99\)00152-2](https://doi.org/10.1016/S0042-6989(99)00152-2)
- 1054 Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in
1055 reading: Comparisons to other models. *Behavioral and brain sciences*, *26*(4), 445–476. <https://doi.org/10.1017/S0140525X03000104>
- 1057 Renninger, L. W., Verghese, P., & Coughlan, J. (2007). Where to look next? eye movements reduce local
1058 uncertainty. *Journal of vision*, *7*(3), 6–6. <https://doi.org/10.1167/7.3.6>
- 1059 Risko, E. F., Stolz, J. A., & Besner, D. (2010). Spatial attention modulates feature crosstalk in visual
1060 word processing. *Attention, Perception and Psychophysics*, *72*(4), 989–998. [https://doi.org/10.](https://doi.org/10.3758/APP.72.4.989)
1061 3758/APP.72.4.989

- 1062 Romani, C., Di Betta, A. M., Tsouknida, E., & Olson, A. (2008). Lexical and nonlexical processing
1063 in developmental dyslexia: A case for different resources and different impairments. *Cognitive*
1064 *neuropsychology*, 25(6), 798–830. <https://doi.org/10.1080/02643290802347183>
- 1065 Romani, C., Ward, J., & Olson, A. (1999). Developmental surface dysgraphia: What is the underlying
1066 cognitive impairment? *The Quarterly Journal of Experimental Psychology Section A*, 52(1), 97–
1067 128. <https://doi.org/10.1080/713755804>
- 1068 Saghiran, A., Valdois, S., & Diard, J. (2020). Simulating length and frequency effects across multiple tasks
1069 with the bayesian braid-phon model. In S. Denison, M. Mack., Y. Xu, & B. Armstrong (Eds.),
1070 *Proceedings of the 42nd annual conference of the cognitive science society*. Cognitive Science
1071 Society.
- 1072 Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems*
1073 *Research*, 1(4), 201–220. [https://doi.org/10.1016/S1389-0417\(00\)00015-2](https://doi.org/10.1016/S1389-0417(00)00015-2)
- 1074 Scarf, D., Boy, K., Reinert, A. U., Devine, J., Güntürkün, O., & Colombo, M. (2016). Orthographic
1075 processing in pigeons (*columba livia*). *Proceedings of the National Academy of Sciences*, 113(40),
1076 11272–11276. <https://doi.org/10.1073/pnas.1607870113>
- 1077 Seidenberg, M. S. (2012). Computational models of reading: Connectionist and dual-route approaches. In
1078 M. Spivey, K. McRae, & M. Joanisse (Eds.), *Cambridge handbook of psycholinguistics* (pp. 186–
1079 203). Cambridge University Press.
- 1080 Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisi-
1081 tion and processing. *Cognitive science*, 23(4), 569–588. [https://doi.org/10.1207/s15516709cog2304_](https://doi.org/10.1207/s15516709cog2304_8)
1082 8
- 1083 Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition
1084 and naming. *Psychological Review*, 96(4), 523–568. <https://doi.org/10.1037/0033-295X.96.4.523>
- 1085 Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cogni-*
1086 *tion*, 55(2), 151–218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- 1087 Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching
1088 hypothesis. *Journal of Experimental Child Psychology*, 72(2), 95–129. [https://doi.org/10.1006/](https://doi.org/10.1006/jecp.1998.2481)
1089 [jecp.1998.2481](https://doi.org/10.1006/jecp.1998.2481)
- 1090 Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of
1091 self-teaching. *Journal of Experimental Child Psychology*, 87(4), 267–298. [https://doi.org/10.](https://doi.org/10.1016/j.jecp.2004.01.001)
1092 [1016/j.jecp.2004.01.001](https://doi.org/10.1016/j.jecp.2004.01.001)
- 1093 Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). Ob1-reader: A model of word recognition
1094 and eye movements in text reading. *Psychological review*, 125(6), 969. [https://doi.org/10.1037/](https://doi.org/10.1037/rev0000119)
1095 [rev0000119](https://doi.org/10.1037/rev0000119)
- 1096 Tucker, R., Castles, A., Laroche, A., & Deacon, H. (2016). The nature of orthographic learning in self-
1097 teaching: Testing the extent of transfer. *Journal of Experimental Child Psychology*, 145, 79–94.
1098 <https://doi.org/10.1016/j.jecp.2015.12.007>

- 1099 Valdois, S., Bidet-Ildei, C., Lassus-Sangosse, D., Reilhac, C., N'guyen-Morel, M.-A., Guinet, E., & Or-
1100 liaguet, J.-P. (2011). A visual processing but no phonological disorder in a child with mixed
1101 dyslexia. *Cortex*, *47*(10), 1197–1218. <https://doi.org/10.1016/j.cortex.2011.05.011>
- 1102 Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological
1103 and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case
1104 studies. *Reading and Writing: An Interdisciplinary Journal*, *16*, 541–572. [https://doi.org/10.](https://doi.org/10.1023/A:1025501406971)
1105 [1023/A:1025501406971](https://doi.org/10.1023/A:1025501406971)
- 1106 Valdois, S., Phénix, T., Fort, M., & Diard, J. (submitted). Atypical viewing position effect in develop-
1107 mental dyslexia : A behavioral and modeling exploration. —.
- 1108 Valdois, S., Reilhac, C., Ginestet, E., & Line Bosse, M. (2021). Varieties of cognitive profiles in poor
1109 readers: Evidence for a vas-impaired subtype. *Journal of Learning Disabilities*, *54*(3), 221–233.
1110 <https://doi.org/10.1177/0022219420961332>
- 1111 Valdois, S., Roulin, J.-L., & Bosse, M.-L. (2019). Visual attention modulates reading acquisition. *Vision*
1112 *Research*, *165*(1), 152–161. <https://doi.org/10.1016/j.visres.2019.10.011>
- 1113 van den Boer, M., van Bergen, E., & de Jong, P. F. (2015). The specific relation of visual attention
1114 span with reading and spelling in Dutch. *Learning and individual differences*, *39*, 141–149. <https://doi.org/10.1016/j.lindif.2015.03.017>
- 1115
- 1116 van den Boer, M., & de Jong, P. F. (2018). Stability of visual attention span performance and its relation
1117 with reading over time. *Scientific Studies of Reading*, *22*(5), 434–441. [https://doi.org/10.1080/](https://doi.org/10.1080/10888438.2018.1472266)
1118 [10888438.2018.1472266](https://doi.org/10.1080/10888438.2018.1472266)
- 1119 van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2013). Modeling the length effect:
1120 Specifying the relation with visual and phonological correlates of reading. *Scientific Studies of*
1121 *Reading*, *17*(4), 243–256. <https://doi.org/10.1080/10888438.2012.683222>
- 1122 Waechter, S., Besner, D., & Stolz, J. A. (2011). Basic processes in reading: Spatial attention as a necessary
1123 preliminary to orthographic and semantic processing. *Visual Cognition*, *19*(2), 171–202. <https://doi.org/10.1080/13506285.2010.517228>
- 1124
- 1125 Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model
1126 and selective literature review. *Psychonomic Bulletin & Review*, *8*(2), 221–43. [https://doi.org/](https://doi.org/10.3758/BF03196158)
1127 [10.3758/BF03196158](https://doi.org/10.3758/BF03196158)
- 1128 Wochna, K. L., & Juhasz, B. J. (2013). Context length and reading novel words: An eye-movement
1129 investigation. *British Journal of Psychology*, *104*(3), 347–363. [https://doi.org/10.1111/j.2044-](https://doi.org/10.1111/j.2044-8295.2012.02127.x)
1130 [8295.2012.02127.x](https://doi.org/10.1111/j.2044-8295.2012.02127.x)
- 1131 Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decod-
1132 ing and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society*
1133 *of London B: Biological Sciences*, *369*(1634), 20120397. <https://doi.org/10.1098/rstb.2012.0397>
- 1134 Zoubinetzky, R., Bielle, F., & Valdois, S. (2014). New insights on developmental dyslexia subtypes:
1135 Heterogeneity of mixed reading profiles. *PLoS ONE*, *9*(6), e99337. [https://doi.org/10.1371/](https://doi.org/10.1371/journal.pone.0099337)
1136 [journal.pone.0099337](https://doi.org/10.1371/journal.pone.0099337)