



**HAL**  
open science

# Visual attention matters during word recognition: A Bayesian modeling approach

Thierry Phénix, Émilie Ginestet, Sylviane Valdois, Julien Diard

► **To cite this version:**

Thierry Phénix, Émilie Ginestet, Sylviane Valdois, Julien Diard. Visual attention matters during word recognition: A Bayesian modeling approach. *Psychonomic Bulletin and Review*, 2025, 10.3758/s13423-024-02591-4 . hal-04874424

**HAL Id: hal-04874424**

**<https://hal.science/hal-04874424v1>**

Submitted on 8 Jan 2025

**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.

# Visual attention matters during word recognition: A Bayesian modeling approach

Thierry Phénix, Émilie Ginestet, Sylviane Valdois, and Julien Diard  
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

It is striking that visual attention, the process by which attentional resources are allocated in the visual field so as to locally enhance visual perception, is a pervasive component of models of eye movements in reading, but is seldom considered in models of isolated word recognition. We describe BRAID, a new Bayesian word Recognition model with Attention, Interference and Dynamics. As most of its predecessors, BRAID incorporates three sensory, perceptual and orthographic knowledge layers together with a lexical membership submodel. Its originality resides in also including three mechanisms that modulate letter identification within strings: an acuity gradient, lateral interference and visual attention. We calibrated the model such that its temporal scale was consistent with behavioral data, and then explored the model's capacity to generalize to other, independent effects. We evaluated the model's capacity to account for the word length effect in lexical decision, for the optimal viewing position effect and for the interaction of crowding and frequency effects in word recognition. We further examined how these effects were modulated by variations in the visual attention distribution. We show that visual attention modulates all three effects and that a narrow distribution of visual attention results in performance patterns that mimic those reported in impaired readers. Overall, the BRAID model could be conceived as a core building block, towards the development of integrated models of reading aloud and eye movement control, or of visual recognition of impaired readers, or any context in which visual attention does matter.


*Keywords:* Bayesian modeling, word recognition, lexical decision, visual attention, optimal viewing position

Reading is a complex phenomenon that draws on a constellation of perceptual and cognitive processes. The complexity of the whole system led to focus on specific aspects of the reading process, so that computational models of eye movement control, word recognition, and reading aloud were mostly independently developed. As the core process that uniquely characterizes reading (Norris, 2013), word recognition is a core component of all specific models whatever

the domain of reading on which they focus (eye movement control or reading aloud). However, because they have been developed to primarily account for specific phenomena in specific tasks, current models differ in the assumptions they feature about word recognition.

Many models have already been designed to specifically account for visual word recognition (for reviews, see Norris (2013), Phénix et al. (2016), and Reichle (2021)). Among the most notable are the *Interactive Activation model* (IA; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), the *Activation Verification* model (AVM; Paap et al., 1982), the *Dual and Multiple Read-Out* models (DROM and MROM; Grainger & Jacobs, 1994, 1996; Jacobs et al., 1998), the *Multiple Object Recognition and attention SElection* model (MORSEL; Mozer & Behrmann, 1990), the *Spatial Coding* model (SOLAR and SCM; Davis, 2010), the *Sequential Encoding Regulated by Inputs to Oscillations within Letter units* model (SERIOL; C. Whitney, 2001; C. Whitney & Lavidor, 2005), the *Open-Bigrams* model (OB and Overlap OB; Grainger & van Heuven, 2003; Grainger et al., 2006), the *Bayesian Reader* model (BR; Norris, 2006; Norris & Kinoshita, 2012; Norris et al., 2010), the *Overlap* model (Gomez et al., 2008), and the *Letter in Time and Retinotopic Space* model (LTRS; Adelman, 2011). Collectively, these

---

The research leading to these results has received funding from the “Fondation de France” (Ph.D. grant to TP). This work includes research conducted as part of Thierry Phénix’s Ph.D. thesis, defended at the Université Grenoble Alpes (UGA) on January, 15th, 2018 (manuscript in French: “Modélisation bayésienne algorithmique de la reconnaissance visuelle de mots et de l’attention visuelle”). Authors wish to thank Pierre Bessi ere and Jacques Droulez for support, inspiration and helpful discussions. Correspondence concerning this paper should be addressed to  Julien Diard, Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105; Universit e Grenoble Alpes, BMD; 1251 Av. des Universit es, CS40700, 38058 Grenoble Cedex 9, France. Email: Julien.Diard@univ-grenoble-alpes.fr

models of isolated word recognition have successfully simulated a wide range of empirical findings from a wide range of experimental tasks. However, they were developed without consideration on how they might contact with models of eye-movement control or reading aloud and they only focused on expert reading, while mostly ignoring constraints from pathological reading or reading development.

Besides, models of reading aloud (Coltheart et al., 2001; Perry et al., 2007, 2010; Plaut et al., 1996; Seidenberg & McClelland, 1989), reading acquisition (Pritchard et al., 2018; Ziegler et al., 2014), and eye movement control (Engbert et al., 2002, 2005; Reichle et al., 1999, 2009) do incorporate a visual word recognition submodel but this submodel is usually less deeply specified than in dedicated models of word recognition. It is further puzzling that, while the role of visual attention in reading is central in models of eye-movement control, no visual attention component is postulated by models of word recognition and reading aloud (with notable exceptions in MORSEL (Mozer & Behrmann, 1990) and the Multi-Trace Model (MTM; Ans et al., 1998), respectively). Recent evidence that the role of visual attention in skilled word processing may have been underestimated (Besner et al., 2016; Lachter et al., 2004; Waechter et al., 2011), alongside evidence that visual attention affects not only visual word processing when impaired (Bosse et al., 2007; Habekost, 2015; Perry & Long, 2022) but also reading acquisition (Franceschini et al., 2012; Gavril et al., 2021; Pasqualotto et al., 2022; Valdois, Roulin, & Bosse, 2019), strongly supports its role in visual word recognition. Therefore, our main goal in the current paper was to develop a new visual word recognition model that capitalizes on previous word recognition models and incorporates visual attention as a key component of single word processing. To fulfill this objective, we introduce the BRAID model (for “Bayesian word Recognition with Attention, Interference and Dynamics”), an original probabilistic and hierarchical model of word recognition.

The rest of this paper is structured as follows. We first conduct an overview of previous word recognition models to define the general, conceptual architecture of the BRAID model. As a synthesis of previous models, BRAID for the first time incorporates all three mechanisms of visual acuity, lateral interference and visual attention. Second, we translate this architecture into a probabilistic model, and show how, from the model, Bayesian inference automatically yields formal expressions for letter identification, word recognition, and lexical decision. Third, we evaluate the model’s capacity to account for key findings in isolated word recognition. We show that the model successfully accounts for context effects on letter perception (including the word and pseudo-word superiority effect) and for the frequency, neighborhood frequency, and transposed-letter priming effects in lexical decision. We then focus on three behavioral effects that more

specifically involve visual and visual attentional mechanisms and show how well the model simulates the length effect in lexical decision, the optimal viewing position, and crowding effects in word recognition. Last, BRAID provides the opportunity to examine how variations in the distribution of visual attention affect word recognition. While previous word recognition models concentrated on fluent reading in expert readers, we show that BRAID performance in a condition of reduced visual-attention distribution parallels the atypical behavior observed in some forms of acquired and developmental dyslexia.

### Theoretical background of the conceptual BRAID model

Since the IA model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), it is assumed that word recognition involves a set of three interacting levels coding for visual features, letters, and words. This “minimal” architecture has been adopted in most models of reading aloud, such as the Dual Route Cascaded model (DRC; Coltheart et al., 2001), the Connectionist Dual Process family of models (CDP, CDP+, CDP++; Perry et al., 2007, 2010, 2014), and the Triangle model (Plaut et al., 1996; Seidenberg & McClelland, 1989). As most of its predecessors, BRAID includes three dynamically evolving and interacting layers, that are the “letter sensory” layer, concerned with extraction of sensory features of letters, the “letter perceptual” layer, concerned with encoding and maintaining letter identity and position, and the “lexical knowledge” layer, concerned with encoding known orthographic words and their relative frequencies. In BRAID, like in IA and most previous models, the flow of information between the letter perceptual layer and the lexical knowledge layer is bi-directional.

Other aspects of word recognition models are less consensual, namely the way letter position is encoded, the nature of the visual mechanisms involved and whether word recognition is modulated by visual attention. These issues are discussed below to justify the theoretical assumptions chosen as ingredients of the BRAID model.

### Letter position encoding

The way letter position is encoded within words is a matter of debate in current word recognition models. The IA assumption of position-specific strict letter encoding was adopted by some subsequent models of word recognition (Norris, 2006) and by most models of reading aloud (Ans et al., 1998; Coltheart et al., 2001; Perry et al., 2007, 2010; Seidenberg & McClelland, 1989). But these models could hardly account for the capacity of human readers to process words in which the order of letters has been changed. A critical issue in subsequent work was to determine how letter position is coded within words. Some models have adopted a relative coding scheme, including SERIOL (C. Whitney, 2001), the open-bigram model (Grainger & van Heuven,

2003; Grainger et al., 2006), and the Local Combination Detector model (LCD; Dehaene et al., 2005). The open-bigram approach successfully accounted for a wide range of empirical findings (Grainger and van Heuven, 2003; Peresotti and Grainger, 1999; Schoonbaert and Grainger, 2004; Snell et al., 2018; C. Whitney et al., 2012; see, however Kinoshita and Norris, 2013) and was used to develop new dual route accounts of orthographic processing (Grainger & Ziegler, 2011).

However, other classes of word-recognition models (see Norris (2013) and Frost (2012) for reviews) assume that the representation of a letter is distributed across adjacent positions in the word letter string (Davis, 2010; Gomez et al., 2008; Kinoshita & Norris, 2009, 2012; Norris & Kinoshita, 2008; Norris et al., 2010). Distributed position-coding models have shown that they could account for a wide range of priming data on letter transposition without postulating any specialized orthographic representation, such as open-bigrams (Kinoshita & Norris, 2013).

The ability of these models to account for most empirical findings lead us to question the open-bigram approach. Four issues have been raised. First, it was argued that the open-bigram models are less parsimonious than the distributed position coding models as they require postulating two levels of orthographic processing: a level of position-specific letter encoding to extract bigrams and a level of position-relative letter encoding corresponding to the open-bigrams themselves (Gomez et al., 2008; Kinoshita & Norris, 2013). Second, while models that assume serial processing, such as SERIOL (C. Whitney, 2001), can account for the generation of open-bigrams, the mechanisms involved in their generation are not made explicit in parallel processing models of word recognition (C. Whitney & Cornelissen, 2008). It was further pointed out that open-bigram coding was well suited to orthographic processing in European languages but could hardly generalize to other orthographic systems, such as Semitic languages in which letter-position coding is less flexible (Frost, 2012). Last, the open-bigram approach is reading-specific while distributed position-coding appears as a more basic property of the visual system that is involved in object recognition more generally (Pelli et al., 2006). BRAID therefore adopts noisy, distributed position coding as a generic property of the visual system that applies to word recognition.

### The visual mechanisms involved in word recognition

Current models further differ in the visual mechanisms involved in letter identity processing. Whereas a visual acuity gradient is a necessary component of models of eye movement control, such as the E-Z Reader (Reichle et al., 2003), the SWIFT model (Engbert et al., 2002, 2005), or the OB1-Reader model (Snell et al., 2018), visual acuity was considered outside the scope of most models of word recognition

and reading aloud except SERIOL (C. Whitney, 2001), SCM (Davis, 2010), and AVM (Paap et al., 1982). Does it mean that the acuity drop-off with eccentricity is not relevant in these latter models? Certainly not. The absence of acuity gradient limits the number of effects they can simulate. Indeed, the acuity drop-off with eccentricity cannot be ignored when focusing on either the length or visual field effect in reading (Veldre et al., 2023; C. Whitney, 2001; C. Whitney & Lavidor, 2005). It is also one of the determinants of the optimal viewing position (OVP) effect, i.e., variations in word recognition depending on where in the word the eyes are fixating (Nazir et al., 1992; Rayner, 1986, 2009). To extend the scope of applicability of the BRAID model compared to extant models, it will feature an acuity gradient.

There is also ample evidence that letter identification within words is affected by crowding (Bouma, 1970; Huckauf and Nazir, 2007; Martelli et al., 2009; Pelli and Tillman, 2007; Pelli et al., 2007; see, for a review, D. Whitney and Levi, 2011). Crowding reduces information on the identity of a target letter when surrounded by other letters. This is thought to arise from a competition between the features of adjacent letters and those of the target (Huckauf et al., 1999; Pelli & Tillman, 2008; Pelli et al., 2004). As a direct consequence, the first and last letters of the word that have only one flanking letter are more easily identified than the inner letters that suffer lateral interference from two flanking letters (e.g., Averbach & Coriell, 1961; Bouma, 1970; Forster, 1976; Humphreys et al., 1990; Mason, 1975; Pitchford et al., 2008; Schoonbaert & Grainger, 2004; Tydgate & Grainger, 2009). More generally, crowding affects the visual identification of objects in clutter (Pelli & Tillman, 2007; D. Whitney & Levi, 2011). So, crowding is not specific to letters (Chicherov et al., 2014) but in particular applies to letters within words.

Given its sensitivity to the physical properties of letter strings, such as eccentricity or inter-letter spacing (Martelli et al., 2009; Pelli et al., 2004), crowding is thought to arise in early visual processing before letter identification (Huckauf & Nazir, 2007; Huckauf et al., 1999). Crowding is further sensitive to higher-level processes. Crowding effects diminishes when string familiarity increases, suggesting interactions with high-level orthographic knowledge (Huckauf & Nazir, 2007). Accordingly, crowding affects reading fluency in skilled readers (Martelli et al., 2009; Pelli et al., 2007) and was reported as stronger in beginning readers and dyslexic individuals (Bouma & Legein, 1977; Callens et al., 2013; Crutch & Warrington, 2009; Kwon et al., 2007; Pernet et al., 2006; Spinelli et al., 2002).

As crowding refers to identity modulation depending on letter position in the string, the ability of word recognition models to account for crowding effects directly follows from their encoding assumptions about letter identity and letter position. In some accounts of the reading system, crowding is

viewed as an independent mechanism that only affects letter identity, as acuity does (Grainger et al., 2016). Such independent coding is assumed in the OB1-Reader (Snell et al., 2018) and SCM (Davis, 2010) models. In these models, letter position encoding within words has no impact on letter identity, so that specific mechanisms that cause weaker activation of the inner than the outer letters are implemented to mimic crowding effects. In contrast, other theoretical frameworks postulate that letter identity uncertainty is just a by-product of letter position coding (Gomez et al., 2008; Norris & Kinoshita, 2012; Norris et al., 2010). In these latter models, position information uncertainty is coded through redistribution of letter identity information over positions, in such a way that letter-identity information leaks to adjacent positions, thus causing letter identity uncertainty at each position.

Beyond the fact that the second class of models offers a more parsimonious account of the crowding phenomenon, they further provide a quite natural explanation of letter identity and letter position uncertainty in assuming that both result from word processing through a noisy channel (Norris & Kinoshita, 2012). In line with these models, BRAID adopts a distributed position coding scheme that causes uncertainty about letter identity and would thus account for crowding effects in word recognition.

### The role of visual attention

Visual attention allows portions of the visual stimulus to be more thoroughly processed than others (M. A. Cohen et al., 2012). Its role is all the more critical when several visual objects are simultaneously available for processing (see, for a review, Carrasco, 2011). Visual attention then acts as a filter that enhances sensory information within the attended region to overcome the limited processing capacity of the visual system. Sentence reading is a particular case involving the processing of multiple visual objects. Models of eye movement control in text reading thus assume that attention allocation plays a critical role (Engbert et al., 2002, 2005; Reichle et al., 1998, 1999, 2003; Snell et al., 2018). In line with behavioral data (Cheal & Gregory, 1997; Deubel & Schneider, 1996; Rayner, 2009; Stolz & McCann, 2000), they posit that covert attentional shifts guide saccades and that allocation of attention to target words is required for their accurate recognition.

Despite evidence that visual attention is allocated to the word(s) to be processed in text reading, models of single word recognition have typically been developed without reference to attention. This may be partly due to the fact that word recognition models are models of expert reading and that word reading in expert readers was viewed as being automatic, therefore attention-free (Brown et al., 2002). However, such a view has been challenged. There is now strong evidence that even skilled readers are unable to identify words at unattended locations, suggesting that attention is a necessary preliminary to visual word recognition (Besner

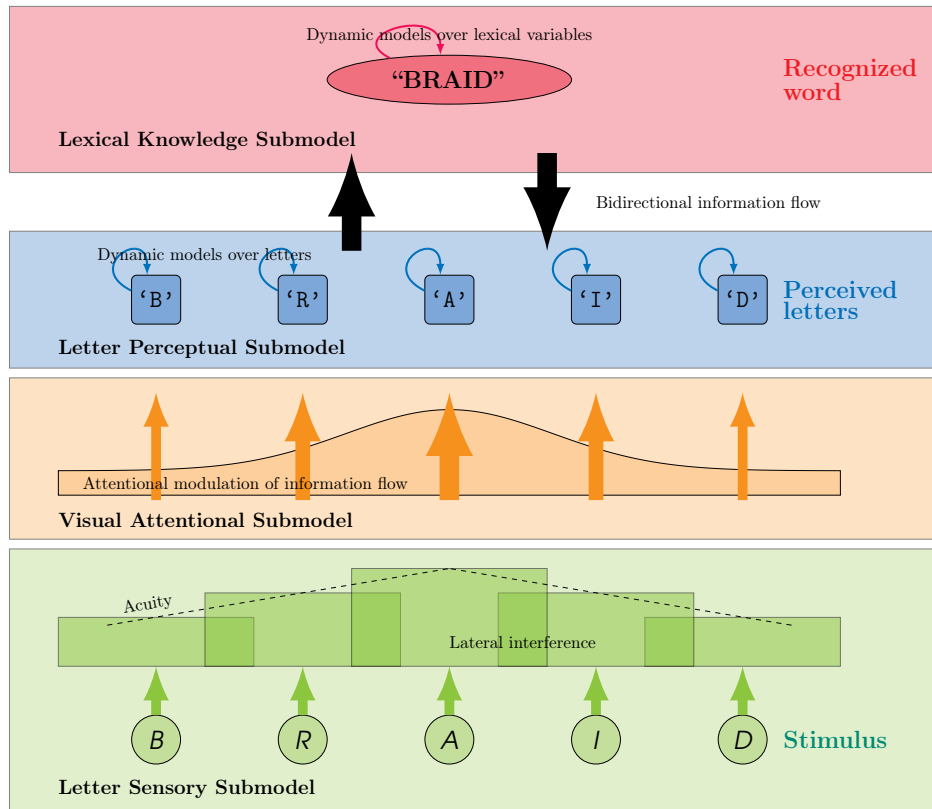
et al., 2016; Lachter et al., 2004; Lien et al., 2010; McCann et al., 1992; Waechter et al., 2011).

Neuropsychological evidence that some forms of acquired and developmental dyslexia follow from a visual attention deficit further argue for visual attention as a key component of the reading system (Bosse et al., 2007; Duncan et al., 1999; Habekost & Bundesen, 2003; Starrfelt et al., 2010; Valdois, 2022). The two MORSEL (Mozer & Behrmann, 1990) and MTM (Ans et al., 1998) computational models that have attempted to account for these forms of dyslexia both include an attentional mechanism that operates at the early stage of visual processing. Disruptions of the attentional device in the two models successfully simulated the atypical word recognition patterns observed in brain-damaged patients. The MTM model further predicted that some forms of developmental dyslexia should result from limited visual attention resources, which was indeed reported (Bosse et al., 2007; Dubois et al., 2010; Lassus-Sangosse et al., 2008; Lobier et al., 2012; Valdois et al., 2003, 2014) and found associated with dysfunction of the dorsal attentional network (Liu et al., 2022; Lobier et al., 2014; Peyrin et al., 2011, 2012; Reilhac et al., 2013; Valdois, Lassus-Sangosse, et al., 2019).

The role of attention was also emphasized in early models of reading acquisition (LaBerge & Brown, 1989; LaBerge & Samuels, 1974), but totally ignored later on (Pritchard et al., 2018; Ziegler et al., 2014), against evidence that visual attention was a concurrent and longitudinal predictor of learning to read (Bosse & Valdois, 2009; Franceschini et al., 2012; Pasqualotto et al., 2022; Valdois, Roulin, & Bosse, 2019; van den Boer et al., 2013; Zoubrinetzky et al., 2019).

To widen the scope of the model, BRAID incorporates a visual attention mechanism as a key component of visual processing in general and reading in particular. Against previous models that confined the role of attention to serial processing (Perry et al., 2007; Vidyasagar, 2005) but in line with MORSEL (Mozer & Behrmann, 1990) and MTM (Ans et al., 1998), we postulate that attention modulates parallel processing of letters during word recognition. In line with behavioral evidence that visual attention interacts with basic visual processes (see for reviews, Carrasco, 2011 or D. Whitney and Levi, 2011), visual attention in BRAID is implemented as an independent mechanism that interacts with visual acuity and lateral interference to modulate letter identification and word recognition.

To summarize, the BRAID model, illustrated in Figure 1, adopts the three representation layers that were implemented in most previous word recognition models, namely the sensory layer (“letter features” level), the perceptual layer (“letter” level), and the lexical orthographic knowledge layer (“word” level). The model assumes parallel letter string processing. Letter identity processing is affected by visual acuity and lateral interference between adjacent letters, which



**Figure 1**

Graphical representation of the structure of the BRAID model, in conceptual form, for the case of a 5-letter input. The three representation layers are the letter sensory layer (green box), the letter perceptual layer (blue box), and the lexical knowledge layer (red box). At the letter sensory layer, visual acuity is represented by decreasing rectangle heights as eccentricity increases (assuming the eye is on the third, central letter), while lateral interference is represented by overlap between adjacent rectangles. The visual attention layer (orange box) inserts an attentional filter between the letter sensory layer and the letter perceptual layer. The arrows width depict the amount of information that is transferred, depending on the position of the letter in the input string (assuming that attention is also focused on the third, central letter). Self-looping arrows (in the blue and red boxes) represent temporal models implementing perceptual accumulation of evidence on letter and word identity. The flow of information between the perceptual and the lexical knowledge layers is bidirectional (black arrows).

further yields positional uncertainty, as in distributed-coding schemes. Finally, an original component of the BRAID architecture is the addition of a visual attentional submodel, allowing to control and spatially modulate the flow of information from the letter sensory submodel to the letter perceptual submodel. Overall, BRAID is the first model of isolated word recognition that incorporates all three mechanism of visual acuity, lateral interference, and visual attention in a single framework.

### Bayesian implementation of the BRAID model

The BRAID model was developed in the probabilistic framework. Probabilistic modeling is a useful tool to the domain of visual word recognition modeling (McClelland, 2013). It is a unified mathematical framework that allows the

definition of hierarchical models and their extension towards complex architectures (Patri et al., 2015), while maintaining the interpretability of the involved representations. In other words, we consider the application of probabilistic modeling at Marr's algorithmic level (Marr, 1982), to define models of cognitive representations and of their hierarchical interactions; we call this "Bayesian algorithmic modeling" (Diard, 2015).

Given the conceptual structure of the BRAID model (Figure 1), we turn to its translation in the probabilistic framework (Bessi re et al., 2013; Diard, 2015; Lebeltel et al., 2004). In this framework, the model takes the form of a joint probability distribution over a set of variables, decomposed as a product of terms. Using conditional independence hypotheses between variables, this product mathematically rep-

resents the architecture we have chosen (see Appendix A). In this section, we highlight some elements of the mathematical definition of the BRAID model and of the probabilistic computations, provided by Bayesian inference, to simulate the cognitive tasks of letter recognition, word recognition, and lexical decision. A graphical representation of the mathematical model is presented on Figure 2 to outline how the different submodels are connected mathematically, and the overall information propagation structure that is entailed. A Python implementation of the BRAID model is available at <https://gricad-gitlab.univ-grenoble-alpes.fr/diardj/braid>.

BRAID is an interactive model with information flows propagating at each iteration from the sensory level to the word level, and from the word level to the perceptual level. Information flow in the model is controlled by coherence variables (Bessi re et al., 2013) represented in Figure 2 by white nodes with  $\lambda$  probabilistic variables. Coherence variables (Bessi re et al., 2013), in a nutshell, can be seen as connecting glue controlling information transfer between the probabilistic variables that they connect. Thus, they can be interpreted as ‘‘Bayesian switches’’ (Gilet et al., 2011), and can be set to one of the three following states: They can be ‘‘closed’’ which allows information transfer, or ‘‘open’’ which blocks information transfer, or, finally, ‘‘controlled’’ (through so-called ‘‘control variables’’) to allow partial transfer of information, in a gradual manner (Ph nix, 2018). When using the BRAID model to simulate the tasks of interest, sensory information ‘‘enters’’ the model as values of the input variables and propagates throughout the model architecture depending on the states of coherence variables. The controlled coherence variables between the sensory and perceptual submodels are particularly relevant here: they are leveraged to mathematically define the model of visual-attention as a filter, to spatially modulate information transfer between the sensory and perceptual submodels.

In the following, we present in general terms the purpose and representational contents of each submodel, providing definitions of probabilistic variables and probability distributions. We introduce the model parameters and their physical interpretation, to directly illustrate the knowledge represented in the submodels. We first focus on bottom-up processing to describe (1) how single letter information is extracted at the sensory level and modulated by the visual similarity between letters, the acuity gradient and interference from adjacent letters, (2) how bottom-up information from the sensory level is accumulated in a representation of the input letter at the perceptual level, and (3) how sensory evidence accumulation is modulated by attention. We then focus on the lexical knowledge submodel which includes a word knowledge component and a lexical membership component. The model assumes bi-directional interactions between the letter perceptual submodel and the word knowledge component. We thus describe (4) the dynamics of let-

ter perception and word recognition over time, and (5) how the evolution of word probability affects letter identification within strings at the perceptual level. Last, we focus on the lexical membership component to describe (6) the dynamics of lexical decision. The complete definition of the model and mathematical derivations that yield the equations we use for task simulations are found elsewhere (Ph nix, 2018).

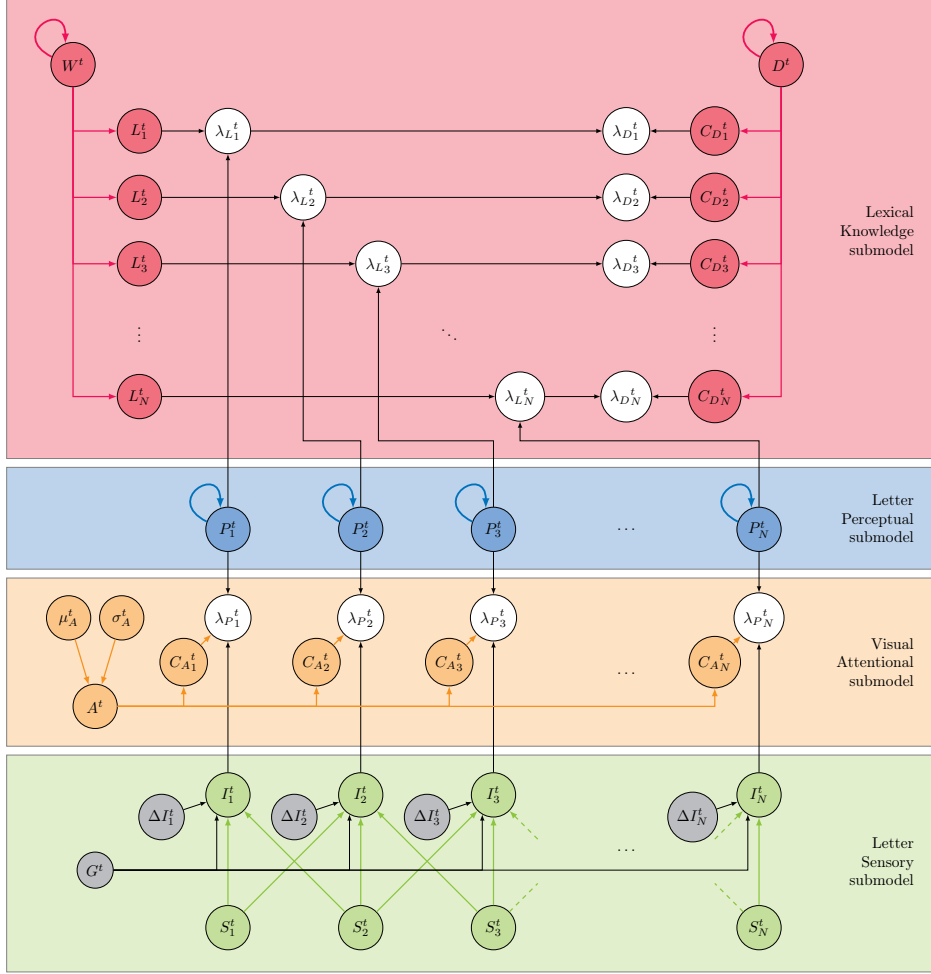
### Letter sensory submodel

The *letter sensory submodel* describes how information is extracted from letter images that are the visual input string, to create sensory representations of letters. Its implementation in BRAID is analogous to the feature level of IA and IA-inspired models (McClelland, 2013; McClelland & Rumelhart, 1981), the ‘‘retina’’ layer of the MORSEL model (Mozer & Behrmann, 1990) or, even more closely relates to the encoding level of the VA model (Paap et al., 1982). Like these, letter identification is modulated by visual similarity between letters. Experimental confusion matrices provide an estimation of inter-letter visual similarity for the Latin alphabet (for reviews, see Mueller and Weidemann, 2012; van der Heijden et al., 1984). A slightly modified version of the Townsend’s confusion matrix (Townsend, 1971) was used in BRAID on the set of letters of the alphabet, enriched by the character ‘‘\$’’, which corresponds to the unknown, illegible, or missing character. We note  $\mathcal{D}_L = \{a, b, \dots, z, \$\}$  this domain of possible characters, of cardinal  $|\mathcal{D}_L| = 27$ .

The original confusion matrix (Townsend, 1971) resulted from the observation of a perceptual process, with a duration that was calibrated so that each subject was, overall, correct on 50% of trials. The duration of this perceptual process was on the order of at least several tens of milliseconds. Overall, our aim was to obtain a temporal granularity, in the model, so that one iteration (from time step  $t$  to  $t + 1$ ) corresponded to one millisecond. Therefore, the sensory confusion matrix was ‘‘scaled down’’ temporally, and this featured an internal parameter ( $Scale_t$ ).

BRAID implements an acuity gradient and lateral interference between neighboring letters as two additional mechanisms that modulate letter identification at the sensory level. The effects of gaze position and visual acuity on letter identification are expressed by the probability distribution  $P(I_n^t | S_{1:N}^t \Delta I_n^t G^t)$ , which represents the probability to identify letter  $I_n^t$  (in position  $n$  at time  $t$ ) given that letters  $S_{1:N}^t$  are presented as stimuli ( $1 : N$  in the subscript is a shorthand to represent the set of variables from position 1 to  $N$ ) and knowing gaze position ( $G^t$ ) over the input letter string at time  $t$ . (the  $\Delta I_n^t$  variable is used technically to track relative positions between the  $I$  and  $S$  variables).

The acuity gradient decreases the quality of sensory processing symmetrically and linearly as the distance from gaze position increases. In probabilistic terms, this is implemented by a decrease of the information content of proba-


**Figure 2**

Graphical representation of the structure of the BRAID model. Nodes represent variables of the model, and arrows relate variables that appear together in a probabilistic term of the joint probability distribution decomposition (nodes with self-looping arrows represent dynamical models where that variable at time  $t$  depends from the same at time  $t - 1$ ). Subscript indexes refer to spatial positions, i.e.,  $X_1, X_2, \dots, X_N$  are positions 1 to  $N$  in the input letter string. Superscript indexes refer to temporal notation, i.e.,  $X^t$  refers to variable  $X$  at time  $t$ .

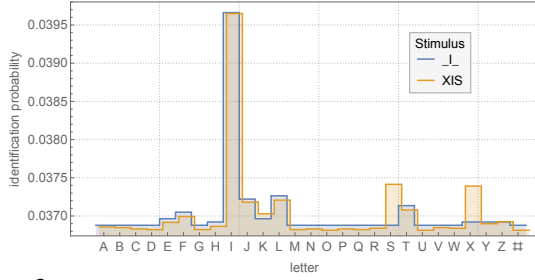
bility distribution  $P(I_n^t | S_{1:N}^t \Delta I_n^t G^t)$ , as a function of eccentricity (the distance between considered position  $n$  and gaze position  $G^t$ ). The strength of this decrease in information content is controlled by parameter  $\theta_G$ : the higher the value of  $\theta_G$ , the less visual acuity modifies the distribution.

When a target letter is processed inside an input letter string, its identification is further affected by interference from neighboring letters. We consider that only adjacent letters influence a given letter, and their relative influence is described by a probability distribution (involving variables  $\Delta I_{1:N}^t$  and a free parameter  $\theta_I$ ). The effect of interference on letter processing is illustrated on Figure 3, which compares the probability distributions at the sensory level for an isolated letter and for the central letter of a triplet. As can

be seen, in the triplet case, competing probability peaks result from the presence of adjacent letters. In the probabilistic framework, because of the normalization rule (probabilities add up to 1), these additional peaks “eat up” a portion of available probability. In other words, any increase in identification probability for a given letter results in a probability decrease for the other letters. This directly follows from probabilistic modeling, mimicking competition between letters at the sensory level.

To summarize, the letter sensory submodel describes how the identity of letters in the word is modulated by (1) the visual similarity of each letter to other letters in the alphabet (through the parameters of the confusion matrix), (2) the distance of the letter from gaze position (through the acuity





**Figure 3**

Illustration of the probability distribution of the identification of the letter input “I” presented in isolation, without lateral interference ( $P(I_n^t | [S_{1:3}^t = \text{“}_I\text{”}])$ , blue curve; and surrounded by two adjacent letters in the string XIS ( $P(I_n^t | [S_{1:3}^t = \text{“XIS”}])$ , orange curve).

gradient parameter), and (3) the presence of adjacent letters (through the lateral interference parameter).

### Letter perceptual submodel: Single letter perception

In the *letter perceptual submodel*, sensory evidence about letter identity accumulates over time to build internal perceptual representations, as in the letter level of the IA model (McClelland & Rumelhart, 1981). This accumulation is performed simultaneously on all positions of the input string.

The perceptual variables  $P_{1:N}^{1:T}$  represent internal letter representations, defined on domain  $\mathcal{D}_L$ . Perceptual variables are dynamic, gradually acquiring, and losing information over time. Before stimulation, the prior distributions over percepts,  $P(P_n^0)$ , define the resting state of knowledge about letters. These priors are defined by uniform distributions over letters. The transition model  $P(P_n^t | P_n^{t-1})$  describes the temporal evolution of letter knowledge, so that perception at time  $t$  is influenced by perception at time  $t-1$ . We assume a gradual loss of perceptual information, so that, in the prolonged absence of sensory information from the input letter string, perceptual information accumulated in  $P_n^t$  would decay back to its resting state (i.e., a uniform distribution); this features a free parameter, noted  $Leak_p$ , which controls the speed of information decay.

We first describe how letter perception proceeds by accumulation of sensory evidence from bottom-up sensory processing, considering the very simple case of an isolated letter presented on gaze fixation as input, i.e., without lexical influence (e.g.,  $s_{1:N}^{1:T} = \text{“}_X\text{”}$ ; in our notation, a capital letter symbol refers to a variable, and a lowercase symbol, here  $s_{1:N}^{1:T}$ , refers to a variable value). Letter recognition amounts to computing the resulting probability distribution over  $P_n^T$ , given the stimulus  $s_{1:N}^{1:T}$  and gaze position  $g^{1:T}$ . This is represented mathematically by computing, in the model, the probability distribution  $P(P_n^T | s_{1:N}^{1:T} g^{1:T} [\lambda_{P_n}^{1:T} = 1])$ , with  $\lambda_{P_n}^{1:T}$

coherence variables to define how information flows in the model (Gilet et al., 2011). We note  $Q_{P_n}^T$  this probability distribution.

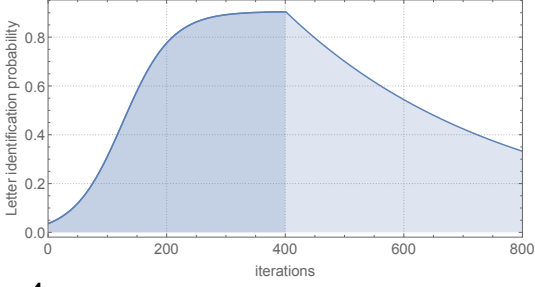
In that simplified case, applying Bayesian inference to the letter sensory submodel and the letter perceptual submodel yields:

$$Q_{P_n}^T = P(P_n^T | s_{1:N}^{1:T} g^{1:T} [\lambda_{P_n}^{1:T} = 1]) \propto \left[ \sum_{P_n^{T-1}} \left[ P(P_n^T | P_n^{T-1}) \times Q_{P_n}^{T-1} \right] \times \sum_{\Delta I_n^T} \left[ P(\Delta I_n^T) \times P([I_n^T = p_n^T] | s_{1:N}^T \Delta I_n^T g^T) \right] \right]. \quad (1)$$

This inference features two elements that are classically found in Hidden Markov Model inference (HMM; Rabiner & Juang, 1993), Bayesian Filtering, or Dynamic Bayesian Network inference (DBN; Murphy, 2002). The first factor (first summation of Equation (1)) corresponds to the prediction term of the inference: it is the product of the recurrence term  $Q_{P_n}^{T-1}$ , that is to say, the result of the same computation at previous time-step  $T-1$ , by the transition term  $P(P_n^T | P_n^{T-1})$ , marginalized over variable  $P_n^{T-1}$ . This first factor is an intermediary result, “halfway between iterations  $T-1$  and  $T$ ” that yields information decay.

The second factor (second summation of Equation (1)) involves the observation or update term which is computed from the letter sensory model. Indeed, it is a weighted sum of sensory letter processing ( $P([I_n^T = p_n^T] | S_{1:N}^T \Delta I_n^T G^T)$  from the letter sensory submodel) over three adjacent positions, with weights being interference strength ( $P(\Delta I_n^T)$ ). This describes how sensory information from the input string is fed into the perceptual submodel, to be accumulated into the perceptual representation of the target letter. Typically, perceptual evidence brings information, reinforcing the perceived letter probability on the correct letter identity hypothesis.

Figure 4 illustrates a simulation of letter identification using Equation (1), in which a single input letter “I” is presented for 400 iterations and then removed. During the first 400 iterations, accumulation of sensory evidence into the perceptual letter representation overcomes information decay, so that the probability distribution over letter  $Q_{P_n}^T$  peaks on the correct hypothesis:  $P([P_n^{400} = \text{“I”}] | S_{1:N}^{1:T} [\lambda_{P_n}^{1:T} = 1] G^{1:T})$  converges, in a sigmoid manner, towards a high value. When the input is removed, only information decay remains, so that probability decreases towards the resting state of the probability distribution, which is uniform. In the above inference, a noteworthy but implicit component is the application of the normalization rule, to output, at each iteration, a proper probability distribution. Probabilistic normalization mimics competition between letters at the perceptual level: it is dictated by the mathematical formalism in the probabilistic framework, and is analogous to explicit lateral inhibition processes in connectionist models. During



**Figure 4**

*Illustration of the time course of the letter identification task, as computed by Equation (1). The plot shows the probability that the letter in the input is an “I”,  $P([P_n^t = \text{“I”}] | S_{1:N}^{1:T} [\lambda_{P_n}^{1:T} = 1] G^{1:T})$  (y-axis), as a function of simulated time  $t$  from 0 to  $T = 800$  iterations (x-axis; with initial probability value at  $1/|\mathcal{D}_L| = 1/27$ ). The input letter “\_” is presented during the first 400 iterations then removed.*

the first 400 iterations, perceptual evidence accumulates in favor of the input letter, which in turn decreases probability for all other competitors. After iteration 400, in the absence of physical input, the probability of the target letter is not increased any more by sensory processing: it suffers from gradual decay and increased competition, until the uniform probability distribution is reached.

We have so far focused on single letter identification to describe how information propagates from the sensory to the letter perceptual submodel, without considering top-down influence from lexical knowledge. However, letter identification within words is a more complex matter that involves both attention and lexical feedback. We start by describing how visual attention is implemented in BRAID and how it modulates letter identification within words.

### Visual attention submodel: Letter perception with attention

The *visual attention submodel* inserts an attentional filtering mechanism between the sensory and perceptual letter submodels, that is to say, a positional modulation of the amount of information transferred from the sensory to the perceptual letter submodel. Its main mathematical feature describes, using a probability distribution, how visual attention is spatially distributed over the input string letters.

Visual attention at time step  $t$  is represented by variable  $A^t$ . Its domain represents discrete positions of letters of the input string. The probability distribution  $P(A^t | \mu_A^t, \sigma_A^t)$  is defined as a Normal distribution of mean  $\mu_A^t$ , which denotes the position of the focus of visual attention, and standard-deviation  $\sigma_A^t$ , which characterizes the spatial dispersion of visual attention (see Figure 5). The total amount of attentional quantity is represented by parameter  $Q_A$  which was kept constant and

equal to 1 in the current paper (for more information, see Steinhilber et al., 2023).

The probability values at each position, given by the probability distribution of attention, control the amount of bottom-up information transfer, from the output of the letter sensory model to the perceptual sensory model for information accumulation (i.e., how much of the probability distribution over  $I_n^t$  is fed to the probability distribution over  $P_n^t$ ).

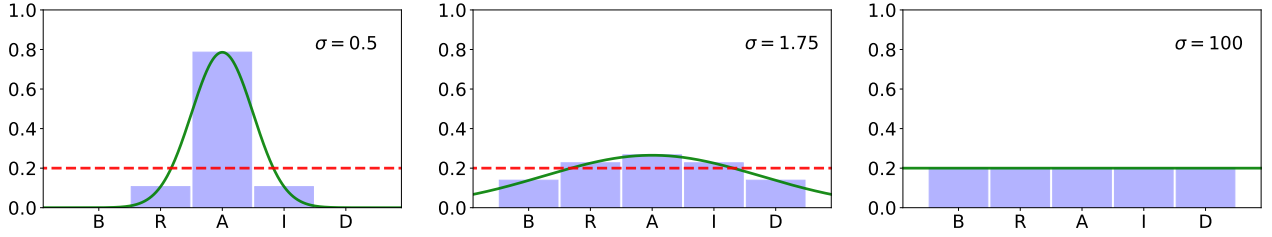
The position  $\mu_A^t$  of attention peak directly follows from the physical constraints of the experimental situation that is to be simulated. For instance, a centrally positioned fixation cross suggests  $\mu_A^t = 3$  for an input letter string of length  $N = 5$ . In the model, attention position  $\mu_A^t$  and gaze position  $g^t$  can be dissociated and set at different positions on the letter string (for an example, see Valdois et al. (2021); however, in the simulations presented below, we only consider cases in which gaze and attention coincide, so that  $\mu_A^t = g^t$ .

The attention distribution for different values of  $\sigma_A^t$  is illustrated in Figure 5. Small values (e.g.,  $\sigma_A^t = 0.5$ ) concentrate attention so that almost only a single letter is efficiently processed. Large values of  $\sigma_A^t$  (e.g.,  $\sigma_A^t = 100$ ) spread attention so that all letters are processed equally, but inefficiently (with only  $1/N$  probability at each position). An intermediate attention dispersion of  $\sigma_A^t = 1.75$ , calibrated from independent data (Ginestet et al., 2019) was used in the experiments as the default  $\sigma_A^t$  value.

We return to our previous inference of isolated letter recognition, to show the mathematical effect of the visual attention distribution. In that case, Equation (1) becomes:

$$Q_{P_n}^T = P(P_n^T | s_{1:N}^{1:T} g^{1:T} \mu_A^{1:T} \sigma_A^{1:T} [\lambda_{P_n}^{1:T} = 1]) \propto \left[ \begin{array}{l} \sum_{P_n^{T-1}} [P(P_n^T | P_n^{T-1}) \times Q_{P_n}^{T-1}] \\ \alpha_n \sum_{\Delta I_n^T} [P(\Delta I_n^T) \times P([I_n^T = p_n^T] | s_{1:N}^T \Delta I_n^T g^T)] \\ \times \left[ \frac{1}{|\mathcal{D}_L|} \right] \end{array} \right], \quad (2)$$

where  $\alpha_n = P([A^T = n] | \mu_A^T, \sigma_A^T)$  is the amount of visual attention allocated to position  $n$  by the visual attention distribution. Compared to Equation (1), the perceptual accumulation of evidence is now the weighted sum of two terms, the first one being the uniform probability distribution (representing lack of information) and the second one being the direct output of the letter sensory submodel. These terms are weighted by attention allocated to the considered position: if all attention is focused to position  $n$ , then  $\alpha_n = 1$ , and all sensory information propagates to percept  $P_n^T$  as previously described (Equation (1)). On the contrary, if no attention is allocated to position  $n$ ,  $\alpha_n = 0$ , then no sensory information propagates to the percept, so that any information previously accumulated in percept  $P_n^T$  would gradually decay over time, as if the input were absent. Any  $\alpha_n$  value between 0 and 1



**Figure 5**

Examples of attention distributions (blue bars) over the input 5-letter string “BRAID”, for a narrow dispersion of attention ( $\sigma_A^t = 0.5$ , left), for a wide dispersion of attention ( $\sigma_A^t = 100$ , right) and for the default situation ( $\sigma_A^t = 1.75$ , center). The red dashed line represents the uniform distribution for a word of this length, as a reference (bars above this reference indicate increased processing for this letter).

thus modulates the amount of sensory information that propagates from the sensory to the perceptual level.

To summarize, letter identity information from the sensory level is modulated by the spatial distribution of visual attention over the input letter sequence. Depending on the location of gaze and attentional focus, and depending on visual attention dispersion, the amount of letter identity information transferred to the letter perceptual submodel differs from one letter to the other. We describe below how letter perception affects the distribution of probabilities for word recognition.

### Lexical knowledge submodel: Modulation of word recognition by attention

In the *lexical knowledge submodel*, words are described as probability distributions over sequences of letters, using a classical naïve Bayes fusion model (Russell & Norvig, 1995), as in the Bayesian Reader model (Norris, 2006; Norris & Kinoshita, 2012). This submodel also describes the dynamic accumulation of perceptual evidence about the identity of the word from a given input letter string, with word frequency as resting state.

Variables  $L_{1:N}^{0:T}$  represent letters that spell known words: for a given word  $w_i$ , for a given position  $n$ ,  $P(L_n^t | [W^t = w_i])$  is a probability distribution over letters that is almost 1 on the correct letter for this word, and a residual,  $\epsilon$  value ( $\epsilon = 10^{-4}$ ) for all other letters.

The lexical submodel further represents the temporal evolution of knowledge about word identity in an input letter string. A dynamical model  $P(W^t | W^{t-1})$  describes the temporal evolution of knowledge about words. Decay speed is controlled by a free parameter  $Leak_W$ . In the absence of input, probability decays towards word frequencies  $P(W^0)$ .

In the complete BRAID model, the visual word recognition task is simulated by computing the probability distribution over words (that we note  $Q_W^T$ ) given sensory stimulation and visual attention parameters; applying Bayesian

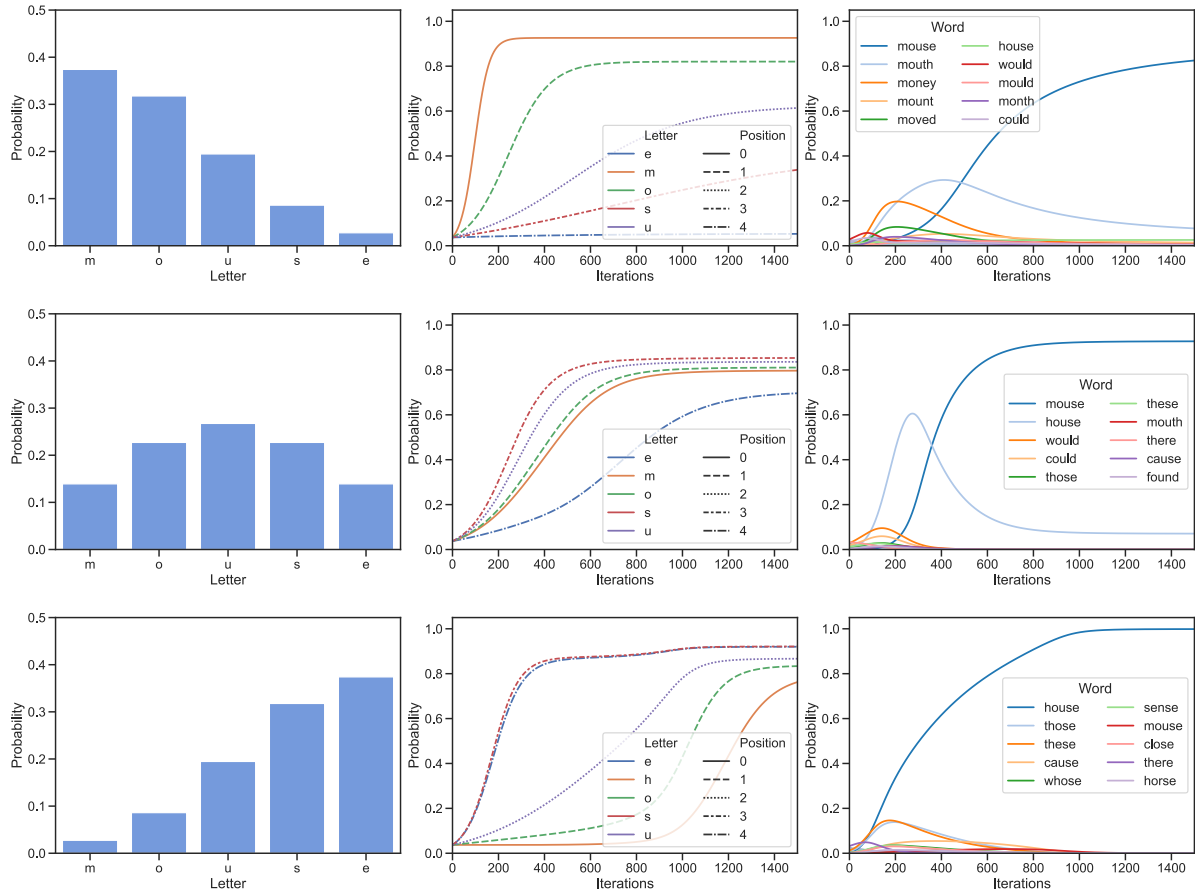
inference to the model yields:

$$Q_W^T = P(W^T | s_{1:N}^{1:T} g^{1:T} \mu_A^{1:T} \sigma_A^{1:T} [\lambda_{L_{1:N}}^{1:T} = 1] [\lambda_{P_{1:N}}^{1:T} = 1]) \propto \left[ \sum_{W^{T-1}} \left[ P(W^T | W^{T-1}) \times Q_W^{T-1} \right] \times \prod_{n=1}^N \langle P(L_n^T | W^T), Q_{P_n}^T \rangle \right], \quad (3)$$

where  $\langle \cdot, \cdot \rangle$  denotes the dot product, and where  $Q_{P_n}^T$  is the simulation of letter identification as described by Equation (2).

Equation (3) features a dynamical evolution of knowledge about word identity, with the recursive term  $Q_W^{T-1}$  combined, under the sum, with the dynamical term  $P(W^T | W^{T-1})$ . This portion of Equation (3) describes gradual decay of information. The rest describes perceptual accumulation of evidence about word identity, with a product over all positions of dot products between the letters “predicted” for a word  $P(L_n^T | W^T)$  and the letters perceived from the input  $Q_{P_n}^T$ . The dot product, here, can be interpreted as measuring the similarity between probability distributions  $P(L_n^T | W^T)$  and  $Q_{P_n}^T$ : its value is high when predicted and perceived letters match, and low when they do not (Steinilber et al., 2022). In the model, this similarity measure implemented by the product of dot products drives word recognition, as it defines the manner in which lexical hypotheses are supported, or not supported, by letter perceptual evidence. This similarity measure is a mathematical consequence of the assumptions on the model architecture (see Figure 2), in the probabilistic setting.

Figure 6 illustrates how visual attention modulates the dynamics of letter and word recognition, considering the input word “MOUSE” as an example. When the focus of attention is on the first letter (Figure 6, top row), identity information accumulates faster for leftmost letters at the perceptual level. As a result, recognition probability increases more rapidly for words sharing their initial letters with the input and/or being of higher frequency. However, over time, more per-



**Figure 6**

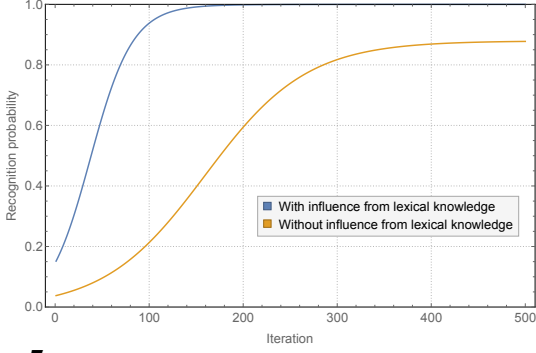
*Illustration of the evolution of letter and word recognition depending on the location of the attention focus (top row: attention on the first letter; middle row: attention on the central letter; bottom row: attention on the last letter) when processing the input letter sequence “MOUSE”. Left plots illustrate the attention distribution assuming default dispersion. Plots in the middle and right columns show the evolution of probabilities (y-axis) for the most likely hypotheses at the letter perceptual level (middle column) and at the lexical level (right column) as a function of iterations from 0 to  $T = 1,500$  (x-axis). Letter position is indicated by the dash type for the curves and letter identity by the line color of each curve (e.g., the long-dash green curve shows the probability that the letter in position 1 is an “O”). At the perceptual level (middle column), only the most likely hypothesis is shown for each position (i.e., not all probability distributions are shown, only the letter hypothesis with the highest probability). At the word knowledge level (right column), only the 10 most likely hypotheses are shown.*

ceptual information about letters accumulates, leading probability for the target word to surpass those of competitors. Note that despite sharing most of its letters with the input and being of higher frequency, the lexical hypothesis “HOUSE” is never “a plausible contender in the probabilistic race”: indeed, in that case, attention is on the first letter, which discriminates between “HOUSE” and “MOUSE”.

The model’s behavior is quite different when attention is centered (Figure 6, middle row). In that case, letter identity information accumulates faster for the inner letters, so that “HOUSE” is the most probable competitor at the beginning of processing. Nevertheless, “MOUSE” wins the race, due

to gradual increase of identity evidence for the initial letter. Finally, consider the situation where the focus of attention is on the last letter of the input word (Figure 6, bottom row). Then, few sensory information is available on the identity of the first letter, so that the competitor that matches letter perceptual evidence and is of higher frequency (i.e., “HOUSE”) is erroneously recognized instead of the target word.

This illustrates how bottom-up information from the perceptual level affects word recognition. We now describe how letter perceptual evidence is modulated by top-down flow of information from the word knowledge submodel.



**Figure 7**

*Time course of the identification of the letter “I”: evolution of the probability that the middle letter is an “I” (y-axis), with the input string “THINK”, with (blue curve) or without (orange curve) influence from lexical knowledge, as a function of simulated time  $t$  from 0 to  $T = 500$  (x-axis).*

### Letter perceptual submodel: Letter in string perception

We have described how letter identification information propagates in the bottom portion of the model, i.e., without lexical influence (see Equation (2)); we now consider letter identification with lexical influence, which we note  $Q'_{P_n}$ .

In that case, applying Bayesian inference to the model to compute  $Q'_{P_n}$  yields:

$$Q'_{P_n} \approx \sum_{W^T} [Q_{W^T} \times P([L_n^T = p] | W^T)] \times Q_{P_n} \quad (4)$$

The model here has the structure of a Hierarchical Hidden Markov Model (HHMM; Murphy, 2002), with two Markov chains evolving in parallel, one over the identity of letters  $P_{1:N}^t$ , the other over word identity  $W^t$ , which inform each other. At each iteration, Equations (2), (3) and (4) are computed in sequence. Since Equation (3) for word recognition involves  $Q_{P_n}$  from Equation (2), this can be interpreted as a “bottom-up” flow of information, from the letter perceptual submodel to the lexical submodel (upward black arrow in Figure 1). Then, since Equation (4) for letter recognition involves  $Q_{W^T}$  from Equation (3), this can be interpreted as “top-down” flow of information, from the lexical submodel to the letter perceptual submodel (downward black arrow in Figure 1). This is repeated for every iteration, until the decision threshold relevant to the task is reached (see Table 1).

Figure 7 illustrates the evolution of recognition probability for target letter “I” in the input string “THINK”, comparing the cases with (i.e., computing  $Q'_{P_n}$  with Equation (4)) and without (i.e., computing  $Q_{P_n}$  with Equation (1)) influence of lexical knowledge. It can be observed that lexical knowledge accelerates recognition of target letter “I”, thus mimicking the well-known word superiority effect.

### Lexical knowledge submodel: Lexical membership

The lexical membership component of the word knowledge submodel determines whether the input letter string is inside or outside the set of known words. This component is useful for performing the lexical decision task, but, importantly, it is not specific to lexical decision. Indeed, we showed that this mechanism could also be used for novelty detection, and was thus critical to trigger novel word orthographic learning during reading acquisition (Ginestet et al., 2022; Steinhilber et al., 2023). Lexical membership is represented by Boolean variables,  $D^{0:T}$ , with value *True* when the input letter string is a known word of the lexicon. A dynamic model over these variables is composed of a prior distribution  $P(D^0)$ , which we set to a uniform, 50/50 probability distribution (when used to mimic lexical decision experimental situations in which nonwords and words are equally present in stimuli) along with a dynamical model  $P(D^t | D^{t-1})$ , which features a free parameter  $Leak_D$ , that controls decay speed towards the uniform, initial state. The three dynamic models over letters, over words, and over lexical membership, evolve in parallel.

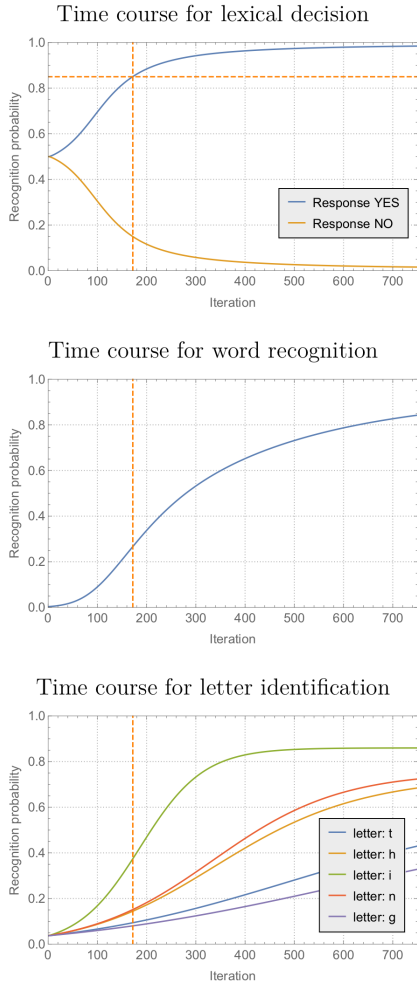
Using Bayesian inference, the lexical membership component allows to dynamically assess whether an input string is a known word or not, by computing a probability distribution noted  $Q_D^T$ . Consider first  $Q_{D=True}^T$ , i.e., the probability that the input is a word. Bayesian inference yields:

$$Q_{D=True}^T = P([D^T = True] | s_{1:N}^{1:T} g^{1:T} \mu_A^{1:T} \sigma_A^{1:T} [\lambda_{D_{1:N}}^{1:T} = 1] [\lambda_{P_{1:N}}^{1:T} = 1]) \propto \left[ \sum_{D^{T-1}} [P([D^T = True] | D^{T-1}) \times Q_{D^{T-1}}] \right] \times \left[ \sum_{W^T} \left[ P(W^T | k^{1:T-1} [\lambda_{L_{1:N}}^{1:T-1} = 1] [\lambda_{P_{1:N}}^{1:T} = 1]) \times \prod_{n=1}^N \langle P(L_n^T | W^T), Q_{P_n} \rangle \right] \right] \quad (5)$$

Consider now the case  $Q_{D=False}^T$ , i.e., the probability that the input is not a word. Bayesian inference yields:

$$Q_{D=False}^T = P([D^T = False] | s_{1:N}^{1:T} g^{1:T} \mu_A^{1:T} \sigma_A^{1:T} [\lambda_{D_{1:N}}^{1:T} = 1] [\lambda_{P_{1:N}}^{1:T} = 1]) \propto \left[ \sum_{D^{T-1}} [P([D^T = False] | D^{T-1}) \times Q_{D^{T-1}}] \right] \times \left[ \sum_{i=1}^N \sum_{W^T} \left[ P \left( W^T \left| \begin{array}{l} k^{1:T-1} \\ [\lambda_{L_i}^{1:T-1} = 0] \\ [\lambda_{L_{1:n}^{1:T-1}} = 1] \\ [\lambda_{P_{1:n}^{1:T-1}} = 1] \end{array} \right. \right) \times \prod_{\substack{n=1 \\ n \neq i}}^N \langle P(L_n^T | W^T), Q_{P_n} \rangle \times \langle P(L_i^T | W^T), (1 - Q_{P_i}^T) / |D_L| \rangle \right] \right] \quad (6)$$

These equations, again, have the same structure as in HMM models, with the first factor featuring the dynamic



**Figure 8**

Comparison of time course in lexical decision (top panel), word recognition (middle panel), and letter identification (bottom panel) for the input letter string “THING”. Each plot shows the evolution of probabilities (y-axis; top:  $Q_{D=\text{True}}^T$  in blue and  $Q_{D=\text{False}}^T$  in orange; middle: most likely hypothesis (word  $W = \text{THING}$ ) in  $Q_W^T$ ; bottom: most likely hypothesis in distribution  $Q_{P_n}^T$  for each position  $n$ ) as a function of simulated time  $t$  from 0 to  $T = 750$  (x-axis). Vertical, dashed lines illustrate what would happen assuming a lexical decision threshold set at .85: lexical decision would proceed even though the probability for the correct word and letters have not yet converged to high probability values.

model and recurrence term, and the second factor featuring the observation model for evidence accumulation. They take a form similar to the answer of visual word recognition of Equation (3), except that, instead of building the probability distribution over words, we marginalize over this distribution, so that all possible words are considered. For instance,

in Equation (5), the probability that the lexical membership variable  $D^T$  is true increases when there is a known word that is probably recognized, or if many words are likely to be recognized (that is, the input, maybe is not a word, but has many orthographic neighbors that are words). In Equation (6), probability that the lexical membership variable  $D^T$  is false increases when there is at least one position  $i$  in which the prediction from the lexicon does not match the input sequence (the  $\langle P(L_i^T | W^T), (1 - Q_{P_i}^T) / |\mathcal{D}_L| \rangle$  term).

Figure 8 illustrates a simulation of lexical decision, word recognition, and letter recognition on the input letter string “THING”. Comparing the dynamics of lexical decision with the other simulated processes shows that lexical decision reaches very high probability values before word recognition and letter recognition do. In other words, the input can be recognized as being a word even though certainty about word identity is not yet reached.

### Model Calibration

The formal definition of the BRAID model includes a number of free parameters listed in Table 1. The lexical parameters are easily defined from available lexical databases; in each of the following experiments, the chosen language and lexicon will be made precise. The time-scaling parameters have been calibrated elsewhere on independent sets of data (Phénix, 2018), with the goal to establish correspondence between simulated time and physical time such that one iteration roughly corresponds to one millisecond. To do so, we have first identified reference times from behavioral studies in the literature. Then, we simulated these behavioral studies while varying the free parameters in a given range of values. We finally selected parameter values for which simulated and observed times correspond.

For instance, consider parameters  $Leak_P$  and  $Scale_I$ , respectively controlling the information decay speed in the probability distribution over perceived letters, and the amount of sensory information fed into these. These two parameters jointly affect the dynamics of the accumulation of perceptual evidence about letters. They have been calibrated such that the probability of the correct letter should reach .95, in 150 iterations on average, which is consistent with neuroimaging studies (Madec et al., 2012; Pykkänen & Marantz, 2003; Tarkiainen et al., 1999). In all calibration experiments, we observe that parameter fit is robust, that is to say, small variations of parameter values generate small and smooth variations of model fit.

### Evaluation of the BRAID model

So far, we have described the BRAID model and emphasized both the features it shares with previous word recognition models and those that make it unique. We now evaluate how well the model accounts for behavioral data. The

**Table 1***Default parameter values of the BRAID model.*

Name	Default value	Interpretation	Determination
<b>Lexical parameters</b>			
$ \mathcal{D}_W $	number of words	Domain size for variables $W^{0:T}$ : lexicon size	calibrated from lexicon
$p_{w_i}$	one value per word	$P(W^0)$ parameters: word frequency	calibrated from lexicon
$\theta_{LW}$	one value per word	Peak letter in $P(L_n^t   W^t)$ distributions: word spelling	calibrated from lexicon
$\epsilon$	$10^{-4}$	Residual probability value of letters in incorrectly spelled words	calibrated empirically
$N$		Input letter-string length $\in \llbracket 2; 15 \rrbracket$	simulation dependent
<b>Time-scaling parameters</b>			
$p_{i,s}$	one value per letter pair $< i_n^t, s_n^t >$	$P(I_n^t   S_{1:N}^t \Delta_n^t G^t)$ parameters, summarizing sensory letter decoding	calibrated (Townsend, 1971)
$Leak_W$	1,250	Parameter for information decay in dynamic model $P(W^t   W^{t-1})$	calibrated (Phénix, 2018)
$Leak_D$	0.15	Parameter for information decay in dynamic model $P(D^t   D^{t-1})$	calibrated (Phénix, 2018)
$Leak_P$	$10^{-4}$	Parameter for information decay in dynamic model $P(P_n^t   P_n^{t-1})$	calibrated (Phénix, 2018)
$Scale_I$	5.8	Parameter for scaling down information content of distribution $P(I_n^t   S_{1:N}^t \Delta_n^t G^t)$	calibrated (Phénix, 2018)
<b>Decision thresholds</b>			
$\tau_p$	0.9	Decision threshold for letter recognition	calibrated (Phénix, 2018)
$\tau_w$	0.9	Decision threshold for word recognition	calibrated (Phénix, 2018)
$\tau_d$	0.9	Decision threshold for lexical decision	calibrated (Ginestet, Phénix, Diard, & Valdois, 2019)
<b>Visual and visuo-attentional parameters</b>			
$\theta_G$	1.0	Parameter for spatially scaling down information of distribution $P(I_n^t   S_{1:N}^t \Delta_n^t G^t)$ : decrease of visual acuity as a function of eccentricity	calibrated (Phénix, 2018)
$\theta_I$	0.675	Parameter controlling lateral interference between letters in distribution $P(I_n^t   S_{1:N}^t \Delta_n^t G^t)$ : $\theta_I$ for letter at position $n$ , $(1-\theta_I)/2$ interference from adjacent letters (renormalized for outside letters)	calibrated (Phénix, 2018)
$Q_A$	1.0	Total amount of attentional quantity	simulation dependent
$g^t$	$(N+1)/2$	Gaze position: by default, set at the input letter-string center	simulation dependent
$\mu_A^t$	$(N+1)/2$	Mean of attention distribution $P(A^t   \mu_A^t \sigma_A^t)$ : by default, set at the input letter-string center	simulation dependent
$\sigma_A^t$	1.75	Standard-deviation of attention distribution $P(A^t   \mu_A^t \sigma_A^t)$	calibrated (Ginestet, Phénix, Diard, & Valdois, 2019)

first critical test for a new model of word recognition is to successfully account for key empirical findings, well captured by previous models. BRAID was evaluated on its capacity to explain the frequency effect (more specifically explored through the Bayesian Reader model (Norris, 2006)), the neighborhood frequency effect (simulated within the IA (Jacobs & Grainger, 1992), Multiple-Read-Out (Grainger & Jacobs, 1996), or Bayesian reader (Norris, 2006) frameworks), the word and pseudo-word superiority effect and more general context familiarity effects on letter perception (well captured by the IA model (Rumelhart & McClelland, 1982), the AV model (Paap et al., 1982), or DROM (Grainger & Jacobs, 1994)), and the transposed-letter priming effects (simulated using the Open-bigram model (Schoonbaert & Grainger, 2004), SCM (Davis, 2010), Overlap (Gomez et al., 2008), SERIOL (C. Whitney & Cornelissen, 2008), or the noisy-position version of the Bayesian Reader (Norris et al.,

2010)). The challenge for BRAID was to simulate this large variety of behavioral effects while systematically using the same default parameters. For each of these effects, we selected leading studies in the field and tested how well BRAID predicted behavioral findings when using exactly the same items as in the experimental studies.

Evidence that BRAID fares well in accounting for the frequency and neighborhood frequency effects has already been reported. With respect to frequency, BRAID successfully simulates the log-frequency effect reported in the Chronolex Megastudy (Ferrand et al., 2011) for lexical decision (Saghiran et al., 2020). The model was further challenged on its capacity to simulate frequency effects in lexical decision and word recognition, using the original method introduced by Norris (2006). Results are reported in Phénix (2018), showing that the performance of BRAID and the Bayesian Reader is similar. These findings and the corresponding code

are available at <https://gricad-gitlab.univ-grenoble-alpes.fr/diardj/braid>. As for the neighborhood effect, Phénix et al. (2018) demonstrated that the model faithfully reproduces neighborhood density and frequency effects in lexical decision, and provides new insights on the apparently contradictory, facilitatory and inhibitory, effects reported in behavioral studies.

To examine more in depth whether the model performs as well as its predecessors, we further checked whether BRAID could account for the (pseudo)word superiority and general context familiarity effects on letter perception, and for transposed-letter priming effects in lexical decision. The results of the simulations are reported in Appendix B and the overall effects predicted by the model are summarized in Table 2.

In a first series of three simulations, we show that BRAID successfully reproduces the behavioral findings reported by Johnston (1978) for the word superiority effect, by Chase and Tallal (1990) for the pseudo-word superiority effect, and by Rumelhart and McClelland (1982) for generalization to context familiarity. Two further simulations were performed to simulate conditions in which the word superiority effect was abolished. In the first, following Johnston and McClelland (1974), letter identification was tested within words and non-words, in two conditions in which attention was either focused on the target position (i.e., cued condition) or deployed over the whole input string. As behaviorally observed, BRAID replicates both the expected word superiority effect in condition of whole-word processing and abolition of the effect in the cued-condition. In the second, following Carr et al. (1976) and Massaro (1973), BRAID predicts the expected word superiority effect (higher letter identification within words than non-words or in isolation) in standard conditions of presentation but the effect was abolished when inter-letter spacing was increased and when attention was focused on the target letter position. Overall, following IA (McClelland & Rumelhart, 1981), BRAID is capable of accounting for the basic word and pseudo-word superiority effects and successfully reproduces the higher perceptibility of target letters when embedded in word-like rather than random consonant strings. Moreover, unlike its predecessors, the model can further predict abolition of the word superiority effect under specific conditions of the visual attention distribution.

Last, we evaluated BRAID's ability to generate priming effects in lexical decision, as a way to study the model's letter position coding assumptions. The model was first challenged on its capacity to predict the behavioral performance reported by Norris et al. (2010) for five letter words in various lexical decision priming conditions with identity, transposed, substituted, repeated, and unrelated primes. BRAID reproduces priming effects of remarkably similar direction and amplitude as in the experiment. Second, we assessed

whether the model could replicate priming effects when using more extreme forms of transposed letter primes that are particularly challenging for word recognition models. For this purpose, we referred to the study carried out by Guerrero and Forster (2008) in which significant priming effects were reported for 8-letter words, even when as few as two target letter positions were preserved in the prime. Results of the simulations showed a good fit to the data. BRAID simulated reliable priming effects when all internal letter-pairs were reversed, or when only the two first or the two last letters were preserved. Overall, these findings suggest that BRAID can account for both basic and extreme letter-transposition priming effects on word recognition.

Thus, BRAID was first tested against the basic benchmark effects that motivated the development of earlier word recognition models. As can be seen from Table 2, BRAID is able to simulate a broad range of experimental findings in letter perception, word recognition, and lexical decision. Furthermore, the model successfully accounts for behavioral data, such as abolition of the word superiority effect in a cued condition or priming effects in condition of extreme letter transposition, that are challenging for any word-recognition model. However, our main purpose in the present paper was to evaluate BRAID's capacity to simulate behavioral effects in which visual acuity, lateral interference, and visual attention are more specifically involved. For this purpose, we now focus on the length effect in lexical decision, the optimal viewing position (OVP) effect, and crowding effects in word recognition. To better understand the role of visual attention, we examine whether and how variations in visual attentional distribution over the input string modulate these effects and whether atypical visual attention distributions result in simulated effects that mimic those reported in some forms of dyslexia.

In the following, for each behavioral effect, we propose a brief description of the behavioral data and how the effect was interpreted in the literature and/or simulated in previous computational models. We then report the results of the simulations conducted with BRAID. The model was systematically evaluated on its capacity to simulate the expected findings while using the model default parameters, thus without any parameter adaptation or ad-hoc modification. The distribution of visual attention was then systematically varied to study how these variations affected the effect under concern.

### **The word length effect in lexical decision**

First, we challenged BRAID to account for the word length effect in lexical decision (LD). In LD tasks, words and pseudo-words are visually displayed and the participant must decide as accurately and as fast as possible whether the displayed item is a word or not. There is consistent evidence that LD times relate to the number of letters within words (for a review see Barton et al. (2014)). Behavioral mega-studies



**Table 2**

Summary table of the benchmark effects simulated by BRAID (for details, see Appendix B). LD: Lexical Decision; WR: Word Recognition; RT: Reaction Time.

Description of the effects	Behavioral data	Simulation reference
<i>Frequency effects</i>		
Shorter RTs for higher than lower frequency words for LD using the Chronolex database.	Ferrand et al. (2011)	(Saghiran et al., 2020)
Shorter RTs for higher than lower frequency words in LD and WR.	Norris (2006)	Phénix (2018)
<i>Opposite neighborhood frequency effects in LD</i>		
Inhibitory neighborhood frequency effect: Shorter RTs for words with higher frequency neighbors.	Perea and Pollatsek (1998)	Phénix et al. (2018)
Facilitatory neighborhood frequency effect: Shorter RTs for words with than without higher frequency neighbors.	Siakaluk et al. (2002)	Phénix et al. (2018)
<i>Context effects on letter identification</i>		
Word superiority effect: Better letter identification when presented within words than in isolation.	Johnston (1978)	Appendix B, Simulation 1
Pseudo-word superiority effect: Better letter identification within pseudo-words than nonwords.	Chase and Tallal (1990)	Appendix B, Simulation 2
Context familiarity effect: Better identification of target letters when embedded in word-like than in random consonant strings.	Rumelhart and McClelland (1982)	Appendix B, Simulation 3
<i>Attention modulation of the word superiority effect</i>		
Abolition of the word superiority effect when target position is predictable.	Johnston and McClelland (1974)	Appendix B, Simulation 4
Abolition of the word superiority effect when letter spacing is increased and attention focused on target location.	Carr et al. (1976); Massaro (1973)	Appendix B, Simulation 5
<i>Transposed-letter priming effects in LD</i>		
Adjacent transposed-letter priming (5-letter words): Shorter RTs when the target word is primed by a transposed-letter.	Norris et al. (2010)	Appendix B, Simulation 6
Extreme transposed-letter priming (8-letter words): Reliable priming effects when only a very few letters occur in target position within the prime.	Guerrera and Forster (2008)	Appendix B, Simulation 7

in healthy participants (Ferrand et al., 2011; New et al., 2006) reported slower LD times for longer words (although the relationship was not always linear). An increase of some milliseconds per additional letter (from 9 to 30 ms/letter (Barton et al., 2014)) was reported for expert readers. Higher word-length effects characterize poor and atypical readers (Barton et al., 2014; Reinhart et al., 2013; Valdois et al., 2003, 2011). Extreme length effects on LD times were even reported in letter-by-letter readers, translating into hundreds of milliseconds (or even, several seconds) per additional letter (Barton et al., 2014).

The length effect is typically interpreted as a failure of whole word parallel processing and over reliance on serial processing (Coltheart et al., 2001; Perry & Ziegler, 2002; C. Whitney, 2018). The simulation of this effect was found challenging for the computational models, like BRAID, that posit letter parallel processing within words (Ans et al., 1998; Coltheart et al., 2001; Perry et al., 2007; Seidenberg & McClelland, 1989). These models typically predicted no length effect for words. Behavioral findings suggest that the effect might originate from visual factors. This is supported by evidence from typical readers showing that the word length effect was larger in tasks that more specifically tapped visual processing (Ferrand et al., 2011) or in conditions where either letter visibility or word format was degraded (Arguin & Bub, 2005; L. Cohen et al., 2008; Fiset et al., 2006). Atypical word length effects were also reported in association with a visual attention deficit in developmental dyslexia (Juphard et al., 2004; Valdois et al., 2003, 2011).

## Method

The English Lexicon Project (ELP; Balota et al., 2004, 2007; New et al., 2006) is a database of LD reaction times (RTs) obtained from expert readers on a large set of 38,483 English words. For the present study, first, we configured the BRAID model with the ELP lexicon, and then we selected as stimuli a subset of 1,050 words ranging from 5 to 11 letters in length (150 words per length). All words were mono- or bi-morphemic and of medium-to-high frequency (mean frequency = 373). The analysis of behavioral RTs on this word-set showed a significant length effect ( $t = 5.35; p < .001$ ), corresponding to RTs increasing by 13.64 ms per additional letter on average. In this sample, the length effect was not modulated by word frequency ( $t = 0.49; p = .62$ ).

We first examined whether the model could account for the behaviorally observed word length effect. For this purpose, in a first simulation, we used a unique central gaze and attentional focus ( $G^t$  and  $\mu'_A$  positioned to the center of the input letter string, whatever its length). The model's default values were used for all the other parameters (see Table 1). This simulation featured word processing within a single fixation (hereafter referred to as the No-Shift condition).

The same default parameters were used in a second simulation but the model was allowed to make an additional attentional fixation on longer words (from 7 to 11 letters), based on behavioral evidence that longer words are typically fixated more than once (Rayner, 2009). For short words (4 to 6 letters), the single attentional fixation was the same as in the No-Shift condition. For long words (7 letters or more) the

first attentional fixation lasted 200 iterations and was positioned to the middle of the first half of the word. The second attentional fixation lasted until the lexical decision threshold was reached and was positioned to the middle of the second half of the word. We refer to this condition as the Shift condition.

In a third simulation, the visuo-attentional dispersion  $\sigma_A$  was set to 0.5 to study whether and how an abnormally limited attention distribution would affect the word length effect on LD times. Attention was positioned, in turn, over each letter, for 100 iterations each, and cycles back over letters if necessary. Such a narrow distribution allowed processing around a single letter at once, thus mimicking a condition of strictly serial processing.

### Results

Word length effects on lexical decision RTs are shown in Figure 9 (a) for the behavioral and the simulated data. BRAID predicts a significant word length effect in all three conditions of no-shift ( $t = 30.28; p < .001$ ), shift ( $t = 10.05; p < .001$ ), and strictly serial processing ( $t = 85.31; p < .001$ ). It further predicts large variations in the amplitude of the word length effect depending on the condition.

In the No-Shift condition, the simulated word length effect was far larger than observed for expert readers (139 iterations/ms per additional letter vs. 13.64 ms for the human data). In the shift condition, simulated RTs were drastically decreased for longer words and a length effect of 14 iterations/ms per additional letter was obtained that is consistent with the length effect reported in the behavioral data (13.64 ms). In the strictly serial reading condition using narrow visual attention distribution, the word length effect was of larger magnitude (48 iterations/ms per additional letter) than for expert readers, but of smaller amplitude than in the No-Shift condition. There was no significant Length-by-Frequency interaction in any of the simulation conditions (all  $ps > .05$ ).

Linear regressions between observed and simulated data are presented on Figure 9 (b), for all the conditions. All three simulations well accounted for the empirical data (all  $R^2 > .80$ ), but parameters of the regression indicate that the Shift condition provided the best fit to the human data. The intercept of 512.3 iterations is consistent with the 400 to 500 ms expected for motor response time, and the slope of the regression is close to 1 (1.015), suggesting that the temporal calibration of the model was adequate.

Overall, BRAID can simulate word length effects in lexical decision. The model successfully accounted for human data, when assuming two attentional fixations for longer words, each fixation capturing roughly half of the stimulus. This “partly serial, partly parallel” visual attention processing yielded faster and more realistic RTs for longer words,

compared with fully parallel processing (i.e., the no-shift condition) or strictly serial processing (narrowed VA distribution) of the entire word letter string. The model predicts that the word length effect in lexical decision can be modulated by two factors: the number of attentional shifts (No-shift vs. Shift) and visual attention dispersion (spread or narrowed).

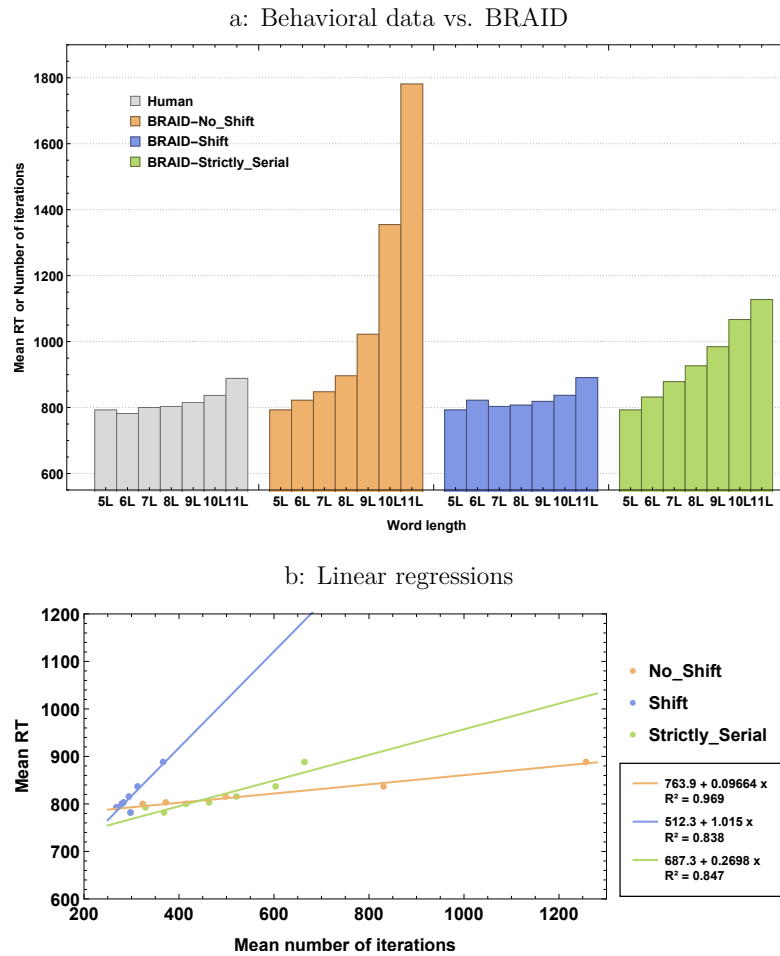
### The optimal viewing position effect

The optimal viewing position (OVP) effect corresponds to the fact that word recognition is more efficient for some specific positions of eye fixation within words. In languages that are read left-to-right, fixating slightly left of the word center allows the optimal recognition of words (Aghababian & Nazir, 2000; Brysbaert & Nazir, 2005; Nazir, 1991; Nazir et al., 1992, 1998; O’Regan & Jacobs, 1984; O’Regan et al., 1984). Performance declines asymmetrically when deviating from the OVP, with an advantage for fixations on the first half of words. As a result, inverted-J-shaped curves characterize word recognition performance as a function of eye position. These asymmetrical viewing position curves have been reported in typical readers (Aghababian & Nazir, 2000; Nazir et al., 1992) but symmetrical inverted V-shaped curves were described in atypical readers (Aghababian & Nazir, 2000; Bellocchi & Ducrot, 2021; Dubois et al., 2007).

The OVP effect was interpreted as potentially resulting from visual, linguistic and/or attentional factors. The acuity drop-off with eccentricity was viewed as a critical factor (Nazir et al., 1991; O’Regan & Jacobs, 1984). Fixating near word center minimizes the loss of acuity as compared to fixating the word outer letters. However, visual acuity cannot be the only factor at play, since the acuity drop-off would predict optimal recognition when fixating at the middle of words, not leftward from the middle. Additional factors related to lexical/orthographic constraints, as higher informativeness of the initial part of words, may further contribute to the OVP effect (Brysbaert & Nazir, 2005; Clark & O’Regan, 1999; Stevens & Grainger, 2003). The effect may also reflect more efficient allocation of attention to the right hemifield (Brysbaert et al., 1996).

### Method

A first series of simulations was carried out by reference to the study of Montant et al. (1998). Since the experiment was conducted with French participants, we configured the model’s lexical knowledge with a reference French lexicon (FLP; Ferrand et al., 2010). The model was presented the same set of 250 5-to-9 letter words (50 per length) as in the experiment. Each word was divided into five zones of equal width and the center of each zone was designated as a potential fixation point. As in the behavioral experiment, simulations were performed by setting gaze position  $G^l$  and attention position  $\mu_A^l$  at the center of the five fixation zones



**Figure 9**

Word length effects in lexical decision. Panel a: Mean behavioral RTs (grey bars) and simulated RTs for the three conditions of No-Shift (i.e., single attentional fixation, orange bars), Shift (i.e., two attentional fixations from 7-letter words, blue bars), and strictly serial reading (green bars), as a function of word length. In the Figure, simulated RTs were scaled and adjusted by aligning them on the behavioral RTs for five-letter words. Other length conditions are predicted by the model. Panel b: linear regressions between simulated and behavioral RTs for the three conditions. Parameters of the regressions ( $y = intercept + slope * x$ ) and  $R^2$  are indicated in the caption box.

for each word length. A grid search method was applied to examine the effect of a large range of variations of the attention distribution parameter ( $\sigma_A$ ) on word recognition. For this purpose, word recognition was simulated for each word length, each of the five fixation zones, each iteration and each  $\sigma_A$  value (ranging from 0.5 to 3) and the mean square error (MSE) between simulated and observed performance was computed.

Then simulated curves were computed for three contrasted levels of visual attention distribution. A  $\sigma_A$  value corresponding to the default value of 1.75 was adopted first to simulate typical OVP effects for each word length. Second, the  $\sigma_A$  parameter was set to 100 to mimic a condition in which

visual attention was maximal and equally spread over letters. Last,  $\sigma_A$  was set to .5 to simulate a narrow distribution of visual attention in which most attention (80%) was allocated to a single letter at a time.

**Results**

Results of the grid search method are provided in Figure 10 together with the simulated word recognition curves for the three conditions of  $\sigma_A$  values and for each word length, as a function of gaze position. Each panel shows the viewing position curve simulated by the model and the curve reported by Montant et al. (1998) for expert readers.

Inspection of the grid search method results (Figure 10,

left column) revealed that the fit between simulated and empirical data was good for a range of  $\sigma_A$  values (the darker zone on the plots corresponds to better fits). The best fit was obtained for a  $\sigma_A$  value of 1 for 5-letter words, 1.5 for 6-letter words, and 2 for words of length 7 or more. The interest of the grid search results is twofold. First, they suggest that the model would capture human data in a robust fashion for a variety of  $\sigma_A$  values. Second, they show that, for all lengths, the region with the best fit to human data (darker regions) always included the model's default value for  $\sigma_A$  (i.e.,  $\sigma_A = 1.75$ ). For the following results, a single "consensus" value (300) was chosen for the number of iterations, as it also provides good fit to the data for all lengths (although, not the best fit for each length).

As shown in Figure 10(second column), using the default  $\sigma_A$  value of 1.75 yielded simulated curves very similar to those reported in expert readers. For all lengths, word recognition was optimal when fixating the word center (or slightly left of center) and the simulated OVP effect was of similar amplitude as for the human data. As expected, word recognition was higher when fixating leftward and the left-right asymmetry of simulated curves increased with word length.

Results of the simulation using a  $\sigma_A$  parameter value of 100 are displayed on Figure 10 (third column). As shown on the figure, the use of a uniform attention deployment over the input word affected both the height and the shape of the viewing position curve. For all word lengths, simulated word recognition probability at the OVP was lower than for expert readers. Furthermore, fixating the rightmost letters (thus, the final part of the word) did not yield the expected strong recognition probability decline, thus resulting in viewing position curves that were less asymmetrical than behaviorally reported in typical readers.

Atypical viewing position curves were also obtained (Figure 10, last column) when the simulation was run using a narrowed attention distribution ( $\sigma_A = 0.5$ ). Then, word recognition was far lower at the OVP for the model than for expert readers and the atypical viewing position curves were characterized by either a reversed-V-shape or an M-shape, depending on word length.

Overall, the BRAID model equipped with its default visual attention parameters was able to successfully account for both the OVP effect and the inverted J-shaped curves typically reported in expert readers. The model further predicted that conditions of uniform or narrowed visual attention distribution would affect not only word recognition efficiency at the OVP but also the shape of viewing position curves for all word lengths. Interestingly, changes in the distribution of visual attention over the word letter-string was sufficient to modulate the OVP and viewing position curve shapes. Variations were obtained while the acuity gradient and lateral interference parameters remained unchanged (set to default values).

## The crowding effect

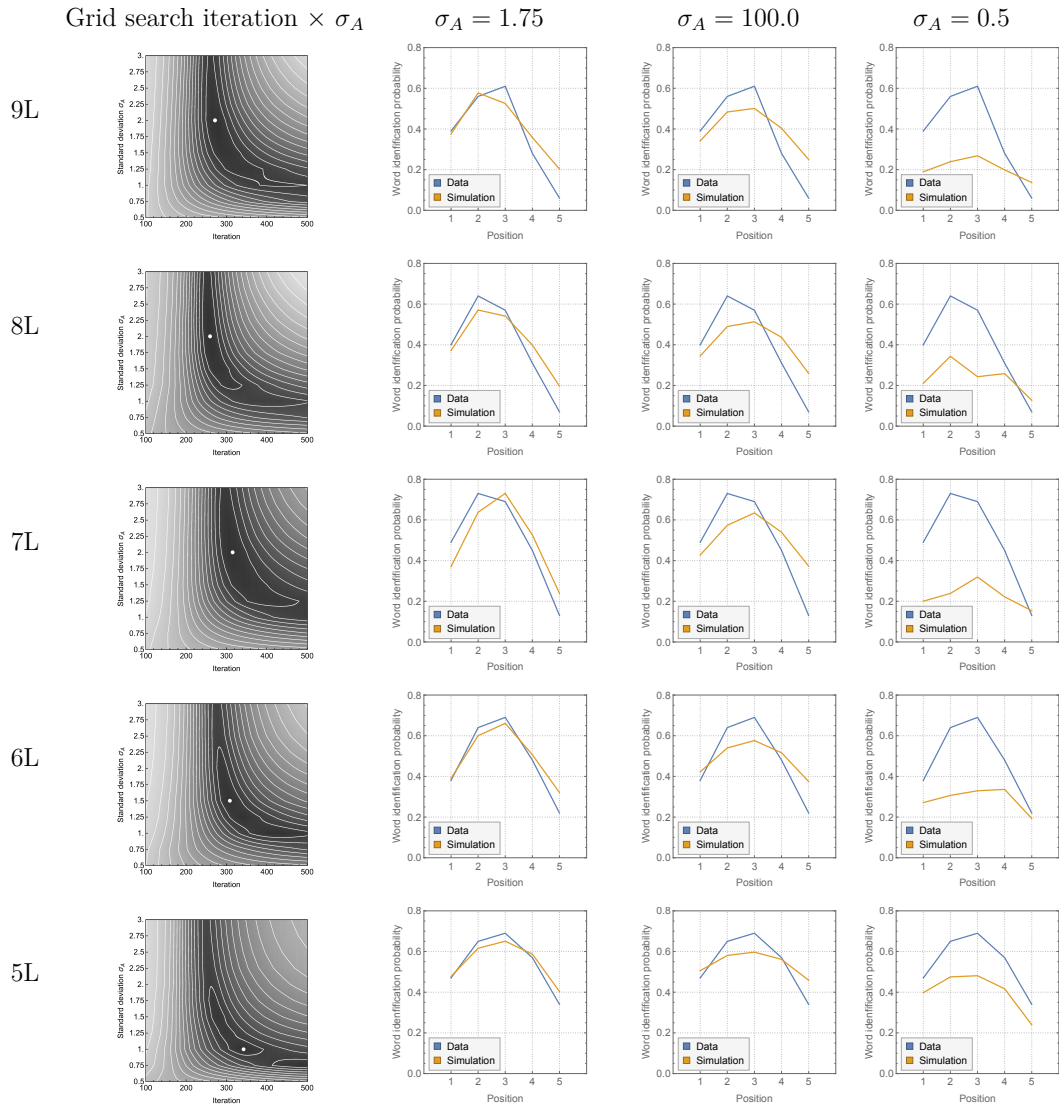
Crowding refers to impaired recognition of a target due to the presence of nearby flankers. It is modulated by the target-flanker spatial distance (Bouma, 1970; Pelli & Tillman, 2007; D. Whitney & Levi, 2011). In reading, a number of studies have explored how variations in inter-letter spacing affected word recognition (Joo et al., 2018; Martelli et al., 2009; Pelli & Tillman, 2007; Perea & Gomez, 2011; Spinelli et al., 2002; Zorzi et al., 2012). They showed that slightly larger than standard spacing resulted in more accurate word recognition and faster processing. However, an increase beyond a critical spacing disrupted word processing (L. Cohen et al., 2008; Spinelli et al., 2002). Not surprisingly, word recognition was adversely affected in conditions of decreased spacing (Montani et al., 2015). Crowding is a major bottleneck for reading. Its effect is stronger in less skilled readers (Kwon et al., 2007) and excessive crowding was reported in developmental dyslexia (Callens et al., 2013; Martelli et al., 2009; Spinelli et al., 2002).

Some findings suggest that crowding interacts with word frequency and word length (however, Perea and Gomez (2011)). Longer words were more prone to suffer from a higher degree of visual crowding, probably due to the combined effects of lateral interference and eccentricity (Martelli et al., 2009; McDonald, 2006). Larger word frequency effect was further reported when the space between letters was greater than normal (Paterson & Jordan, 2010).

Crowding was beyond the scope of most word recognition and reading models. However, in their computational model of word recognition and eye movement control, Snell et al. (2018) assumed that letter identification within words was modulated by visual acuity and visuo-spatial attention. The model further included a "crowding" parameter which differently affected letter identification depending on letter position within words. In the present study, we used BRAID as an experimental substitute to predict the complex interrelations between inter-letter spacing, frequency, and length in word recognition. The model was challenged on its capacity to simulate crowding effects without any dedicated mechanism or additional parameter. The effect of visual attention on crowding was studied through simulations performed first in a model with the default Gaussian distribution of visual attention then, in a model with a uniform distribution.

## Method

For this experiment, the model's lexical knowledge was configured in English with the BLP lexicon (Keuleers et al., 2012). The simulations were conducted using two sets of 50 5-letter words and 50 9-letter words. The words were arbitrarily assigned contrasted frequency values 5, 10, 50, 100, 500, 1,000 to generate series of 6 identical couples of 5- and 9-letter words that only differed in frequency.



**Figure 10**

Results of the simulations on the Optimal Viewing Position (OVP) effect. Each row concerns words of different lengths, from 9-to-5 letter words (from “9L” to “5L”). The “Grid search” column shows contour plots of the mean square error between simulated and human data as a function of simulation iterations and variance  $\sigma_A$  of the attention distribution (in grayscale, with dark zones corresponding to better fits). For each word length, the white dot indicates the best-fit parameter value for  $\sigma_A$  and the number of iterations. The three last columns show plots of the simulated (orange curves) and human (blue curve) OVP effect (word identification probability, y-axis, curves as a function of gaze position, x-axis) for three conditions of visual attention distribution: the default distribution ( $\sigma_A = 1.75$ ); a uniform distribution ( $\sigma_A = 100$ ), and a narrow distribution ( $\sigma_A = .5$ ).

We simulated word recognition for each frequency value while varying inter-letter spacing. For this purpose, both the acuity gradient  $\theta_G$  and lateral interference  $\theta_l$  parameter values were manipulated, with a systematic grid search method. The visual acuity gradient parameter  $\theta_G$  was varied from 1 (its default value) to 5 (i.e., strong acuity decline), with steps of 1. As compared to the default value of  $\theta_l = 0.675$  (thus,

$1 - \theta_l = 0.325$ ), the interference strength  $1 - \theta_l$  was varied from 0.5 (i.e., exaggerated lateral interference) to 0 (no interference from adjacent letters), with steps of .05. The effect of these variations on word recognition RTs was first evaluated using the default visual attention distribution value ( $\sigma_A = 1.75$ ), then a uniform distribution ( $\sigma_A = 100$ ).

## Results

The specific and combined effects of lateral interference and acuity modulation on RTs are presented for short and long words as a function of frequency and visual attention distribution in Figure 11. Word recognition RTs were analyzed with a linear mixed effects model with items as crossed random effects and lateral interference, crowding, frequency, length, and attention distribution as fixed effects.

As shown on Panels (a) and (b) of Figure 11, whatever word length and frequency, the model predicts that words are recognized faster when they suffer either lesser lateral interference ( $[F(10, 64353) = 3159, p < .0001]$ ) or lesser acuity drop-off ( $[F(4, 64353) = 4992, p < .0001]$ ).

Note that these two effects are not independent. In real experimental conditions, lesser lateral interference due to larger inter-letter spacing increases the physical length of the word letter string. In turn, higher physical length increases performance drop-off due to acuity. In the model, the two combined effects do result in a trade-off ( $[F(40, 64353) = 3.051, p < .0001]$ ), whereby letter spacing is both beneficial to word recognition due to lateral interference decrease, and detrimental to word recognition due to greater acuity decline. As a result, see Figure 12 (red trajectory), a slight increase of physical inter-letter spacing translates into lower lateral interference ( $1 - \theta_l$  goes to 0) while only minimally affecting acuity ( $\theta_G$ ). In this condition, word recognition is accelerated. However, beyond a certain  $1 - \theta_l$  value, larger spacing does not compensate for excessive acuity decline ( $\theta_G$  increases when  $1 - \theta_l$  decreases), which negatively affects word recognition. Following such a trajectory in parameter space, the model predicts first a decrease followed by a steady increase in recognition times, consistent with behavioral observations (L. Cohen et al., 2008; Spinelli et al., 2002).

As shown on Figure 11, in the condition of default attention distribution, the model predicted both a word frequency and a word length effect on recognition times. First, the combined effects of lateral interference and acuity gradient were modulated by word frequency ( $[F(275, 64353) = 4752.07, p < .0001]$ ). Stronger lateral interference and lesser visual acuity were more detrimental for word recognition when the word was of lower frequency. Reversely, recognition times at the optimal spacing (the lower part of the red curve) were only slightly affected by variations in word frequency. (The model further predicted minimal effects of frequency in the extreme conditions of simulations characterized by either no interference but maximal acuity, or strong interference but poor acuity. Note that these two conditions are purely theoretical but highly unrealistic physically.)

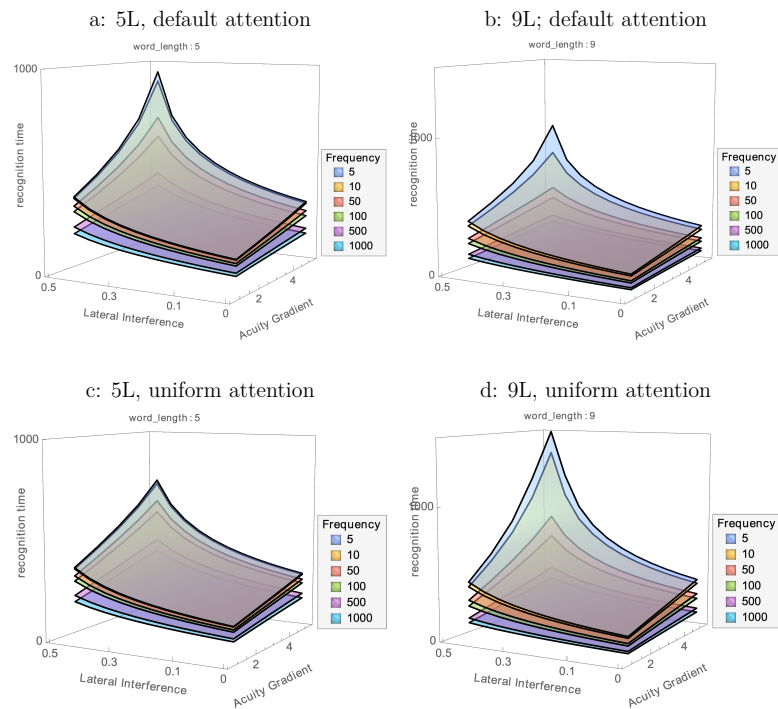
Simulations further revealed that the lateral interference and acuity gradient effects were exacerbated by word length ( $[F(54, 64353) = 58.67, p < .0001]$ ). Longer words were more affected than shorter words by either an increase in lateral interference or a decline in visual acuity, and the length

effect was of higher amplitude for words of lower frequency ( $[F(275, 64353) = 236.08, p < .0001]$ ). As a result, the combined opposite effects of lateral interference and acuity gradient were more pronounced for words of lower frequency and higher length. The model thus predicts that the positive effect of lesser lateral interference on word recognition would turn negative faster for longer words of lower frequency, due to stronger decline in visual acuity and higher competition with more frequent lexical neighbors. The same main and combined effects of lateral interference, acuity gradient, length, and frequency were replicated in the condition of uniform attention distribution.

Comparison of panels (a) and (c) for 5-letter words and panels (b) and (d) for 9-letter words allowed exploring the impact of visual attention distribution on crowding effects. Simulated data showed that the lateral interference by acuity gradient interaction was modulated by attention distribution ( $[F(54, 64353) = 123.20, p < .0001]$ ). Interestingly, opposite effects of attention distribution were observed for the 5- and 9-letter words ( $[F(55, 64353) = 261.52, p < .0001]$ ). For 5-letter words, the effects of lateral interference and acuity gradient were far reduced in condition of uniform attention distribution (Figure 11, panel (c)) relative to the Gaussian attention distribution condition (Figure 11, panel (a)). In contrast, stronger effects of lateral interference and acuity gradient were generated for 9-letter words in the uniform attention condition.

These opposite effects illustrate how the modulation of letter identification by visual attention affects word processing depending on letter spacing. In the uniform attention distribution, the total amount of attention available for processing was equally distributed over the letter string, so that more attention was allocated to each letter in 5- than in 9-letter words ( $1/5$  vs.  $1/9$  respectively). This conferred an advantage to shorter words in the uniform attention condition, compared to the Gaussian distribution. In the uniform condition, enough attention was allocated to each letter of the 5-letter words for their accurate and fast identification. As a result, word recognition was less affected by the negative impact of stronger lateral interference and acuity decline in this condition than in the Gaussian distribution condition, in which the letters that received a lesser amount of attention were more slowly identified. However, the same uniform attention distribution had an opposite effect on 9-letter words. Indeed, the small amount of attention allocated to each letter then slowed down identity information accumulation in each position, so that word recognition for 9-letter words was more strongly affected by the detrimental effects of lateral interference and acuity decline in the uniform than in the Gaussian distribution condition.

Therefore, overall, the model accurately predicts transitory positive effects of inter-letter spacing as far as lateral interference can be reduced enough without affecting signif-



**Figure 11**

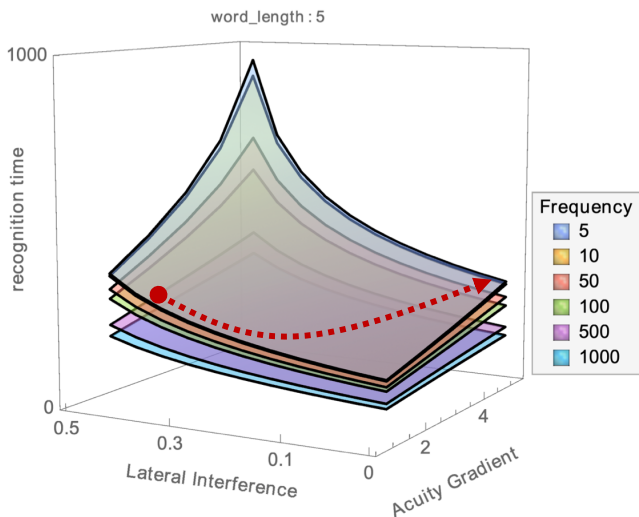
*Effects of variations in lateral interference and visual acuity on RTS for long and short words according to their frequency and visual attention distribution. Top panels (a) and (b) illustrate results using the default attention parameter value, and bottom panels (c) and (d), using a uniform attention distribution. Left-column panels (a and c) are for 5-letter words, right-column panels (b and d) for 9-letter words. Each panel presents the average simulated RTs (z-axes; note that panels a and c go up to 1,000 and panels b and d go up to 1,500) as a function of lateral interference parameter  $1 - \theta_1$  (x-axes, 0 represents no interference), acuity gradient parameter (y-axes, 0 represents no acuity drop-off, 5 represents a strong acuity drop-off), and word frequency (colored curves).*

icantly the drop of visual acuity. However, the initial facilitator effect of increased letter spacing on word recognition quickly turned inhibitory due to stronger acuity decline. The detrimental effect of acuity decline on word recognition was partly compensated by word frequency but exacerbated when the word was longer. More importantly for the present purpose, the model further predicts that the effects of lateral interference and acuity gradient on letter identification and word recognition would be modulated by visual attention. The negative effect of acuity drop off was partly compensated when each one of the word letter received enough attention, which boosted word recognition. In contrast, when only a small amount of attention was allocated to each of the word letters, then letter identification and word recognition were both slowed-down, all the more than the word was longer and the acuity decline stronger.

### Discussion

We have described BRAID, a new model of single word recognition whose originality resides in the implementation

of four mechanisms that modulate letter identification within strings – namely, letter confusability, acuity decline, lateral interference, and visual attention. Simulations of three empirical phenomena – namely, the word length effect in lexical decision, the OVP effect, and the crowding effect in word recognition – were then performed to more specifically challenge the visual and visual attention mechanisms that are implemented in BRAID and better understand how these mechanisms interact between each other and with lexical knowledge. The overall simulation results showed that the model successfully accounted for all three empirical effects, and furthermore, that good fit to the empirical data was not bound to dedicated parameter values specific of each effect but could follow from the use of a single set of independently fixed default parameter values. The same default parameters were also used to account for the classical effects of context on letter perception and frequency, neighborhood frequency and priming in lexical decision. This opens up the prospect of a single stable model able to account for a large variety of tasks and behavioral phenomena. To better



**Figure 12**

*Illustration of a plausible trajectory in parameter space resulting from physical manipulation of inter-letter spacing (red curved arrow, starting from default values of parameters).*

understand how visual attention modulated performance, the distribution of visual attention was systematically varied and simulations rerun using either a uniform distribution or a narrowed distribution of visual attention, instead of the default Gaussian distribution. We show that the length, the OVP, and the crowding effects were all strongly affected by changes in visual attention distribution, suggesting that visual attention does matter in word recognition.

### The visual mechanisms of word recognition

In line with its predecessors (Coltheart et al., 2001; Davis, 2010; McClelland & Rumelhart, 1981; Perry et al., 2007, 2010; Reichle et al., 2003, 2009; C. Whitney, 2001), BRAID incorporates a letter confusion matrix but, unlike most previous models, it postulates that three mechanisms of acuity gradient, lateral interference, and visual attention are further involved in letter identification within words. First, like very few models of word recognition (C. Whitney, 2001) but all models of eye movement control in reading (Engbert et al., 2002, 2005; Reichle et al., 1999, 2003, 2009; Snell et al., 2018), BRAID implements a visual acuity gradient. In BRAID, this visual mechanism is expressed mathematically independently from both the confusion matrix – against AVM (Paap et al., 1982) – and letter position encoding – against SCM (Davis, 2010). Implementation of the acuity gradient was critical to successfully simulate the effects of word length, OVP, and increased inter-letter spacing.

Second, in line with several previous models of word

recognition (Davis, 2010; Gomez et al., 2008; Norris et al., 2010), BRAID implements a mechanism of lateral interference for distributed coding of letter position within words. In BRAID as in these other models, letter position uncertainty has for direct consequence to mix the features of letters, so that their identification is more or less accurate (and fast) depending on their position within the word string. However, unlike these other models, BRAID postulates that ambiguity in position and identity is restricted to adjacent letters. The two mechanisms of acuity gradient and lateral interference naturally allowed the model to simulate crowding effects on word recognition.

Last, the main originality of BRAID is the assumption that visual attention is a key component of word recognition. The implementation of a visual attention mechanism is rather unusual in word recognition models but quite common in models of eye movement control in reading (Engbert et al., 2002, 2005; Reichle et al., 1999, 2003, 2009; Snell et al., 2018). In contrast to previous modeling (whatever the class of reading models), visual attention in BRAID was defined by three parameters, namely attention focus location, attention dispersion, and attention quantity. Although the two first parameters alone were manipulated in the present study, the additional impact of attention quantity on word recognition was recently investigated (Steinhilber et al., 2023), suggesting that attention dispersion would be modulated by the quantity of visual attention available for processing.

However, the visual attention submodel implemented in BRAID is incomplete. The model predicts that attention shifting is required to account for the length effect in lexical decision but the shifting mechanisms were under-specified, with fixed positions and fixation duration as a first approximation. A more realistic model of attention has been proposed by Ginestet et al. (2022) and Steinhilber et al. (2023) to account for orthographic learning. In these studies, an attentional exploration mechanism optimizes the gain of information about letter identity over time. This new model predicts the locus of the next attentional fixation and its dispersion in an online manner, thus allowing the flexible adaptation of attention to the word properties. This augmented visual attention submodel provides a reasonable account of eye-movement patterns during single word processing (Ginestet et al., 2022; Steinhilber et al., 2023). This suggests that implementing both a dynamical model of visual attention and the mechanisms of acuity gradient and lateral interference in a single framework might contribute to fill the gap between models of word recognition and eye movement control in reading. Until now, the model was limited to single word processing (for attempts to fill this gap at the sentence level, see Snell et al. (2018) and Veldre et al. (2020)). We anticipate that BRAID's properties might be applied effectively to the reading of connected words, provided the implementation of additional mechanisms of word predictability. A complete



visual attention model interacting online with word properties (e.g., frequency, neighborhood, length, semantic knowledge) would translate into great exploration flexibility. We would therefore expect visual attention to distribute sometimes over several words at a time, as postulated in SWIFT (Engbert et al., 2005) or OB1-Reader (Snell et al., 2018), sometimes over just one word at a time, as postulated in E-Z Reader (Reichle et al., 1999, 2009) or even over only portions of words. However, simulations of the length, OVP, and crowding effects have shown that visual attention hardly compensates for the sharp acuity decline with eccentricity increase. We thus anticipate that decline in visual acuity would impose major constraints on the number of words that can be simultaneously processed.

The four mechanisms postulated by the model as being involved in letter identification within words differ not only in their function but further in their intrinsic properties. The confusion matrix is specific to the language under concern, while the three other mechanisms relate to general properties of the visual system. Among these visual mechanisms, two relate to the physical properties of the stimulus (i.e., the spatial distance between gaze position and the letter, or between letters within words). As a result, letter confusability, visual acuity, and lateral interference are largely independent from the reader's cognitive abilities. Obviously, the status of visual attention is different. Visual attention is a highly flexible mechanism that can adapt to the constraints of the task and word properties. It further adapts to the reader's purpose, which makes it a cognitive mechanism of great significance for word recognition. The implementation of visual attention as a specific submodel in BRAID translates its cognitive status while all the other three mechanisms are implemented at the sensory level of processing. The implementation of all four mechanisms in BRAID provides insights on their delicate interplay, which is discussed in the next section for each of the three effects of word length, OVP, and crowding.

### Simulations of behavioral data and theoretical insights

The main contribution of a novel computational model is to provide new insights on the mechanisms responsible for the behavioral effects that the model can successfully simulate. With this respect, BRAID represents a significant advancement in allowing to better understand the interplay between visual acuity, lateral interference, and visual attention and how these mechanisms interact with lexical constraints to generate the behavioral effects of word length, OVP, and crowding.

**The length effect.** With respect to the word length effect in lexical decision, simulated findings suggest that the word length effect cannot be viewed as a marker of serial processing (Barton et al., 2014). Indeed, an exaggerated word length effect was generated by BRAID in condition of fully parallel processing (i.e., the no-shift condition using default visual

attention parameters). In this condition of single attentional fixation on the word center, an increasing number of letters received lesser attention and suffered from more severe visual acuity decline in longer than in shorter words. These two mechanisms thus combined to more drastically slow down letter identity information accumulation during longer word processing. Slower letter identification yielded weaker lexical activation, thus far longer RTs in lexical decision.

Rather counter-intuitively and against general belief, the word length effect was reduced rather than increased in the serial condition of attentional shifting. Interestingly, the same exact two mechanisms are responsible for the weaker, and more realistic length effect generated in the serial condition. Then, visual attention was spread over about half of the word letters at each fixation, so that more attention was allocated to each letter. In reducing the number of letters that were processed in parallel, attentional shifting further reduced eccentricity, so that the attended letters also suffered from lesser acuity loss. Thus, the two mechanisms, that concurrently impacted letter identification in fully parallel processing, now contributed to enhance letter identification. In this latter condition, BRAID successfully accounted for the LD word length effect reported for humans in the English Lexicon Project (see Ginestet et al. (2019) for similar findings using a dataset from the French Lexicon Project).

Overall, the main theoretical contribution of BRAID on the word length effect is twofold. First, the model warns against straightforward conclusions about the serial or parallel nature of processing, based on the presence or absence of a length effect. A similar warning was previously formulated by C. Whitney (2018). She showed, within the SERIOL model framework, that serial processing does not necessarily translate into length effects in word processing. As a corollary, we show within the BRAID framework that parallel processing is not necessarily associated with the absence of length effects. Another main contribution of BRAID is to show that visual acuity and visual attention both contribute to word length effect in LD. It was already assumed that length effects resulted from less efficient letter information encoding and lexical activation, due to reduced quality of visual input with eccentricity (O'Regan & Jacobs, 1984). In contrast, visual attention and the number of attentional fixations were rarely considered as potential contributors. Current simulation results are at least compatible with previously reported data indicating that visual attention was involved in word lexical decision (McCann et al., 1992) and that attention interacts with word length in perceptual identification tasks (Aucclair & Siéroff, 2002). They are also in line with evidence for exaggerated word length effect in lexical decision in the context of poor visual attention span in developmental dyslexia (Juphard et al., 2004).

**The OVP effect.** Importantly, the same BRAID model without any modification to its parameter values successfully

simulated the OVP effect and viewing position curves depending on word length. The model predicts that the OVP effect on word recognition is mainly determined by the same two visual factors, namely the acuity gradient and visual attention distribution over the word letter string, while being further modulated by lexical constraints. The role of acuity limitations on the OVP effect is well-documented and largely consensual (Nazir et al., 1992; O'Regan & Jacobs, 1984; O'Regan et al., 1984). What remains debated is the origin of the observed asymmetry of the OVP curves. Indeed, while the acuity gradient by itself can explain better recognition when fixating around the word center, the symmetrical decline of acuity with eccentricity cannot account, in isolation, for the observed left-right asymmetry. Beyond acuity effects, the influence of lexical constraints – the number of words that can be accurately processed from a limited amount of sensory information – was proposed to account for the OVP asymmetry (O'Regan & Jacobs, 1984), assuming that words are more easily recognized based on their initial than their final letters (Adelman et al., 2010; Scaltritti & Balota, 2013). The combined effects of acuity drop-off and lexical constraints are highlighted within the BRAID framework when simulations were carried out in condition of uniform attention distribution. The model does predict a left-right asymmetry of the OVP curves due to lexical constraints, but the asymmetry is small and only minimally affected by word length, against the observed empirical findings. In line with previous evidence (Brysbaert et al., 1996; Holmes & O'Regan, 1987; Stevens & Grainger, 2003), this suggests that lexical constraints alone cannot account for the observed asymmetry.

Moreover, the main contribution of BRAID is to predict that the OVP curve shapes are highly sensitive to the distribution of visual attention over the word letter string. Contrary to the uniform distribution condition, the default Gaussian distribution of attention allowed simulating asymmetrical curve shapes that closely matched the observed curves. When fixating the word's first letter, most attention was allocated to the initial letters which boosted their identification and facilitated word recognition. In contrast, focusing visual attention on the word's final letters enhanced their identification but had no facilitating effect on word recognition. The model thus predicts that the observed asymmetry of the OVP curves would result from the impact of visual attention distribution on letter identification and how letter identification affects lexical processing. Note that a potential effect of visual attention in the OVP asymmetry has been frequently suspected and discussed (Aghababian & Nazir, 2000; Bellocchi & Ducrot, 2021; Brysbaert & Nazir, 2005; Ducrot & Grainger, 2007; C. Whitney, 2001) but the potential role of visual attention on the OVP effect was not previously modeled and not investigated in-depth through behavioral studies.

**The crowding effect.** The same model without modification of its parameters of visual acuity and visual attention was also able to successfully simulate the crowding effects reported in humans for words varying in length and frequency. On the one hand, BRAID simulations replicated the empirical findings that word recognition is slowed when letters are more closely spaced than normal, accelerated when interletter spacing is slightly larger than normal (Bouma, 1970; L. Cohen et al., 2008; Martelli et al., 2009; Perea et al., 2011; Spinelli et al., 2002), and slowed down again in condition of extra-large spacing (L. Cohen et al., 2008; Spinelli et al., 2002; Yu et al., 2007).

A first contribution of the model was to reveal how this pattern follows from the interaction between lateral interference and visual acuity. A general prediction of the model is that word recognition is facilitated when lateral interference between adjacent letters is reduced, thus in condition of larger interletter spacing. As the parameter of lateral interference relates to identity and position coding uncertainty, the model quite straightforwardly predicts higher identity leakage and poorer letter position coding, in crowded conditions. However, the positive effect of lateral interference reduction in condition of larger interletter spacing is counter-balanced by the increase of visual acuity drop-off with eccentricity. Combined variations of lateral interference reduction and visual acuity decline provided a good qualitative account of the opposite beneficial and deleterious effects of interletter spacing on word recognition. Beneficial effects of larger interletter spacing would be observed as far as eccentricity is constrained enough to limit the detrimental effects of visual acuity decline. Even though the adopted scaling of lateral interference and visual acuity was arbitrarily defined in the simulations, BRAID well accounts for the observed qualitative pattern of performance.

In addition, the model suggests that searching for a unique optimal acuity-interference combination that guarantees more efficient reading may be misleading, due to additional interactions with word frequency and word length, and further influence of visual attention. Indeed, simulated results in condition of larger spacing suggest that high frequency words would be more resilient than low-frequency words to the detrimental effects of visual acuity, and that longer word recognition would suffer more from acuity decline. These predictions are well in line with behavioral findings that increasing interletter spacing differently affects word recognition depending on word length (L. Cohen et al., 2008; Risko et al., 2011) and that the recognition of higher frequency words suffers less from increased spacing (Huckauf & Nazir, 2007; Paterson & Jordan, 2010; Perea et al., 2011). Current findings clearly suggest that there is no single interletter spacing larger than standard that is optimal to significantly improve reading performance when systematically applied to all the words in a sentence or a text. Although a

fixed and slight interletter spacing may enhance letter identification and boost word recognition to some extent, more efficient word recognition would require adjusting the width of interletter spacing depending on word length and frequency.

In the context of crowding effects, the main theoretical contribution of BRAID is to provide new insights on how visual attention modulates the combined effects of lateral interference and visual acuity. In conditions of standard interletter spacing and for five letter words, a Gaussian distribution of attention allows allocating enough attention to each letter to compensate for the detrimental effects of lateral interference and visual acuity, which boosts word recognition. However, when interletter spacing is increased, the number of letters that receive enough attention for their accurate identification is gradually reduced while the adverse effect of visual acuity increases rapidly (O'Regan et al., 1984). As a direct consequence, larger spacing becomes detrimental as soon as the amount of attention allocated to each letter does not compensate enough for the loss of acuity induced by spacing. It is noteworthy that exactly the same mechanisms account for the differential spacing effects depending on word length. More generally, when the word is physically longer, either due to larger interletter spacing or to a higher number of letters, then attention is less likely to compensate for the drop of visual acuity.

Overall, the model predicts that the distribution of attention strongly constrains the effects of crowding on word recognition. This prediction connects with experimental evidence that visual attention modulates crowding effects in reading (Y. He & Legge, 2017; Risko et al., 2011; Yeshurun & Rashal, 2010). In particular, simulated results suggest that crowding effects would be exacerbated when each word letter receives little attention, which was here simulated by distributing a “normal” quantity of total attention resources uniformly on physically longer words.

### *Visual attention as a key component of word recognition*

The results of current simulations provide strong evidence that visual attention modulates all three effects of word length, OVP, and crowding. In all cases, word recognition depends on the amount of visual attention distributed over the letter string and to what extent the positive effect of visual attention on letter identification compensates for the deleterious effects of visual acuity and lateral interference. Thus, a default distribution of visual attention favored shorter word processing in lexical decision, generating a far stronger length effect than behaviorally observed, so that an attentional shift alone allowed processing longer words as humans do. The OVP location within words and viewing position curve shapes fitted those reported for humans depending on word length when using the same default visual attention distribution. Word recognition was facilitated when the peak of attention was aligned with the initial word letters

that are also the most informative for lexical discrimination but word recognition was degraded when most attention was allocated to the less informative final letters. In the same way, the acuity-visual attention cost-benefit ratio on word recognition was in favor of visual attention when interletter spacing was slightly increased but, up to a certain threshold (or, critical spacing), the ratio was reversed and attention could no longer compensate for the acuity loss.

The model further predicts that a narrow visual attention distribution should result in exaggerated word length effect and atypical viewing position curves (i.e., simulations using a  $\sigma_A$  value of 0.5). These predictions align with behavioral evidence for higher word length effects in beginning readers (Martens & de Jong, 2006; van den Boer et al., 2015; Zoccolotti et al., 2005, 2008) who otherwise exhibit lower visual attention spans than more advanced readers (Bosse & Valdois, 2009). Higher length effects in word recognition and lexical decision have also been reported in dyslexic readers who showed a visual attention span deficit (Juphard et al., 2004; Valdois et al., 2003, 2011). However, the model goes beyond current findings in postulating that the relationship between visual attention and word length effect may be causal, which should be deeply investigated in future research.

With respect to the OVP effect, the very atypical viewing position curves generated by the model in condition of limited attentional distribution have similar shapes to those reported in the few studies that were conducted in dyslexic individuals (Aghababian & Nazir, 2000; Dubois et al., 2007). Atypical curves have sometimes been reported to co-occur with a visual attention deficit in dyslexia (Dubois et al., 2007; Montant et al., 1998; Valdois et al., 2021) but a causal relationship remains to be established. In support of causality, BRAID was able to successfully account for the very atypical curves reported in dyslexic children with a visual attention span deficit when the distribution of visual attention was narrowed and the peak of attention decorrelated from gaze position (Valdois et al., 2021). Interestingly, since a similar narrowing of visual attention distribution translated in both a higher length effect and atypical viewing position curves in current simulations, the model predicts that exaggerated length effects should co-occur with atypical performance in the OVP task in dyslexic individuals. This co-occurrence that was reported only once in a case of letter-by-letter reading (Montant et al., 1998) should be more systematically investigated together with its relationship with visual attention skills.

The predictions with respect to crowding are less straightforward. We showed through simulations that the crowding effect was exacerbated when a limited amount of attention was allocated to each letter within the word (i.e., in the condition of uniform distribution on long words). This suggests that enhanced visual crowding should be observed in typi-

cal readers when attention is distributed over a large array of visual elements in parallel. But spreading an abnormally limited amount of visual attention over the whole letter string (thus, in condition of brief presentation and parallel processing) might also enhance visual crowding, even for shorter words. In contrast, BRAID predicts that the use of a narrow distribution of visual attention concentrating attention on each individual letter (or a few letters) successively might be quite efficient to compensate for crowding effects on letter identification but at the cost of a strong length effect. These predictions may lead to reconsider interpretations of the exacerbated crowding effects reported in dyslexic individuals (Callens et al., 2013; Martelli et al., 2009; Spinelli et al., 2002; Zorzi et al., 2012). In line with previous claims (Bacigalupo & Luck, 2015; Dayan & Solomon, 2010; S. He et al., 1996; Strasburger, 2005), the current findings suggest that a comprehensive account of crowding effects in typical and atypical readers would require considering interactions with the task attentional load and the participant's visual attention skills.

### **Conclusion**

We have described BRAID, a new model of word recognition that for the first time implements mechanisms of visual acuity, lateral interference, and visual attention. BRAID is also the first to account for all three behavioral effects of word length in lexical decision, OVP, and crowding in word recognition. The model places heavy emphasis on the role of attention in word recognition. It offers a highly sophisticated account of visual attention based on the three parameters of attentional focus, attention distribution, and attentional capacity. The implementation of visual attention in a word recognition model opens the perspective to provide a more integrated account of the reading process. Assuming that each attentional shifting would translate to gaze position changes, the model has the potential to predict both cognitive effects in word recognition and oculomotor patterns in single word processing. We showed through simulations that visual attentional shifting yielded serial word processing. This suggests that serial versus parallel processing in the model might directly follow from the distribution of visual attention over the word letter string. The model might then naturally account for the transfer from more serial to more parallel word processing during reading development. Interestingly, contrary to dual route models, there is no dichotomy between serial and parallel processing within the BRAID framework. The two processing modes are reinterpreted along a continuum. Serial processing is not conceived as strictly serial but rather reflects the sequential parallel processing of word parts, thus resulting in a “partly serial, partly parallel” processing mode. Within this framework, reading develops from the serial processing of small word parts (at the extreme, limited to each single letter) to the serial pro-

cessing of larger and larger word parts (at the extreme, extended to the whole word letter string). This predicts a very gradual transition from more serial to more parallel processing during reading acquisition without requiring additional hypotheses on the automation of the serial processing mechanism. Another important contribution of the current model is to demonstrate that changes in visual attention distribution affect word recognition in a way that is similar to what is described in atypical readers. This suggests that, contrary to current computational models of word recognition that focus on fluent reading in expert readers, BRAID provides a parsimonious theoretical framework for further understanding how reading develops and how that development might be impaired.

### **Funding**

The research leading to these results has received funding from the “Fondation de France” (Ph.D. grant to TP).

### **Conflicts of interest / Competing interests**

The authors have no financial or proprietary interests in any material discussed in this article.

### **Ethics approval**

Not applicable

### **Consent to participate**

Not applicable

### **Consent for publication**

All co-authors have been included in the manuscript and consent for publication.

### **Availability of data and materials**

None of the data for the simulation experiments reported here is available, as they were conducted with a Wolfram Mathematica© implementation of the model (a language that is proprietary to Wolfram Research, Inc., Champaign, Illinois, USA).

### **Code availability**

A publicly available Python implementation of the BRAID model was developed later on and is available under the CeCILL-2.1 license, at <https://gricad-gitlab.univ-grenoble-alpes.fr/diardj/braid>.

## References

- Adelman, J. S. (2011). Letters in time and retinotopic space. *Psychological Review*, *118*(4), 570–82.
- Adelman, J. S., Johnson, R., McCormick, S., McKague, M., Kinoshita, S., Bowers, J., Perry, J., Lupker, K., S.J. and Forster, Cortese, M., Scaltritti, M., Aschenbrenner, A., Coane, J., White, L., Yap, M., Davis, C., Kim, J., & Davis, C. (2014). A behavioral database for masked form priming. *Behavior Research Methods*, *46*, 1052–1067.
- Adelman, J. S., Marquis, S. J., & Sabatos-DeVito, M. G. (2010). Letters in words are read simultaneously, not in left-to-right sequence. *Psychological Science*, *21*(12), 1799–1801.
- Aghababian, V., & Nazir, T. A. (2000). Developing normal reading skills: Aspects of the visual processes underlying word recognition. *Journal of Experimental Child Psychology*, *76*, 123–150.
- Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, *105*(4), 678–723.
- Arguin, M., & Bub, D. (2005). Parallel processing blocked by letter similarity in letter-by-letter dyslexia: A replication. *Cognitive Neuropsychology*, *22*(5), 589–602.
- Auclair, L., & Siéroff, E. (2002). Attentional cueing effect in the identification of words and pseudowords of different length. *The Quarterly Journal of Experimental Psychology*, *55A*(2), 445–463.
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell Labs Technical Journal*, *40*(1), 309–328.
- Bacigalupo, F., & Luck, S. J. (2015). The allocation of attention and working memory in visual crowding. *Journal of Cognitive Neuroscience*, *27*(6), 1180–1193.
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.
- Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, *4*(2), 207–228.
- Barton, J. J., Hanif, H. M., Björnström, L. E., & Hills, C. (2014). The word length effect in reading: A review. *Cognitive Neuropsychology*, *5-6*(31), 378–412.
- Bellocchi, S., & Ducrot, S. (2021). Same, same but different: The optimal viewing position effect in developmental dyslexia, developmental coordination disorders and comorbid disorders. *Dyslexia*, *27*(3), 294–311.
- Besner, D., Risko, E. F., Stolz, J. A., White, D., Reynolds, M., O'Malley, S., & Robidoux, S. (2016). Varieties of attention: Their roles in visual word identification. *Current Directions in Psychological Science*, *25*(3), 162–168.
- Bessièrè, P., Mazer, E., Ahuactzin, J. M., & Mekhnacha, K. (2013). *Bayesian programming*. CRC Press.
- Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*(2), 198–230.
- Bosse, M.-L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: A cross-sectional study. *Journal of Research in Reading*, *32*(2), 230–253.
- Bouma, H., & Legein, C. P. (1977). Foveal and parafoveal recognition of letters and words by dyslexics and by average readers. *Neuropsychologia*, *15*(1), 69–80.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177–178.
- Brown, T. L., Joneleit, K., Robinson, C. S., & Brown, C. R. (2002). Automaticity in reading and the Stroop task: Testing the limits of involuntary word processing. *The American Journal of Psychology*, *115*(4), 515–543.
- Brybaert, M., & Nazir, T. (2005). Visual constraints in written word recognition: Evidence from the optimal viewing position. *Journal of Research in Reading*, *28*(3), 216–228.
- Brybaert, M., Vitu, F., & Schroyens, W. (1996). The right visual field advantage and the optimal viewing position effect: On the relation between foveal and parafoveal word recognition. *Neuropsychology*, *10*, 385–395.
- Callens, M., Whitney, C., Tops, W., & Brybaert, M. (2013). No deficiency in left-to-right processing of words in dyslexia but evidence for enhanced visual crowding. *The Quarterly Journal of Experimental Psychology*, *66*(9), 1803–1817.
- Carr, T. H., Davidson, B. J., & Hawkins, H. L. (1978). Perceptual flexibility in word recognition: Strategies affect orthographic computation but not lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 674–690.
- Carr, T. H., Lehmkuhle, S. W., Kottas, B., Astor-Stetson, E. C., & Arnold, D. (1976). Target position and practice in the identification of letters in varying contexts: A word superiority effect. *Perception & Psychophysics*, *19*(5), 412–416.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525.

- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11, 63–65.
- Chase, C. H., & Tallal, P. (1990). A developmental, interactive activation model of the word superiority effect. *Journal of Experimental Child Psychology*, 49(3), 448–487.
- Cheal, M., & Gregory, M. (1997). Evidence of limited capacity and noise reduction with single-element displays in the location-cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1), 51–71.
- Chicherov, V., Plomp, G., & Herzog, M. H. (2014). Neural correlates of visual crowding. *Neuroimage*, 93, 23–31.
- Clark, J. J., & O'Regan, J. K. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research*, 39, 843–857.
- Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contributions of the dorsal and ventral visual pathways. *NeuroImage*, 40(1), 353–366.
- Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, 16(8), 411–417.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.
- Crutch, S. J., & Warrington, E. K. (2009). The relationship between visual crowding and letter confusability: Towards an understanding of dyslexia in posterior cortical atrophy. *Cognitive Neuropsychology*, 26(5), 471–498.
- Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, 117(3), 713.
- Dayan, P., & Solomon, J. A. (2010). Selective Bayes: Attentional load and crowding. *Vision Research*, 50, 2248–2260.
- Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, 9(7), 335–341.
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12), 1827–1837.
- Diard, J. (2015). *Bayesian algorithmic modeling in cognitive science* [Habilitation à Diriger des Recherches (HDR)]. Université Grenoble Alpes.
- Dubois, M., Kyllingsbæk, S., Prado, C., Musca, S. C., Peifer, E., Lassus-Sangosse, D., & Valdois, S. (2010). Fractionating the multi-character processing deficit in developmental dyslexia: Evidence from two case studies. *Cortex*, 46, 717–738.
- Dubois, M., Lafaye de Micheaux, P., Noël, M.-P., & Valdois, S. (2007). Preorthographical constraints on visual word recognition: Evidence from a case study of developmental surface dyslexia. *Cognitive Neuropsychology*, 24(6), 623–660.
- Ducrot, S., & Grainger, J. (2007). Deployment of spatial attention to words in central and peripheral vision. *Perception and Psychophysics*, 69(4), 578–590.
- Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Chavda, S., & Shibuya, H. (1999). Systematic analysis of deficits in visual attention. *Journal of Experimental Psychology: General*, 128(4), 450–478.
- Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42, 621–636.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813.
- Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in psychology*, 2, 306.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42(2), 488–496.
- Fiset, S., Arguin, M., Bub, D., Humphreys, G. W., & Ridoch, M. J. (2006). An attempt to simulate letter-by-letter dyslexia in normal readers. *Brain and Language*, 98, 251–263.
- Forster, K. I. (1976). Accessing the mental lexicon. In R.J.Wales & E.Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). North-Holland.
- Forster, K. I., Davis, C., Schoknecht, C., & Carter, N. (1987). Masked priming with graphemically related forms: Repetition or partial activation. *Quarterly Journal of Experimental Psychology*, A(39), 211–251.
- Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Fa-coetti, A. (2012). A causal link between visual spatial attention and reading acquisition. *Current Biology*, 22, 814–819.
- Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(5), 263–279.
- Gavril, L., Rosan, A., & Szamoskozi, S. (2021). The role of visual-spatial attention in reading development: A meta-analysis. *Cognitive Neuropsychology*, 38(6), 387–407.

- Gilet, E., Diard, J., & Bessière, P. (2011). Bayesian action-perception computational model: Interaction of production and recognition of cursive letters. *PLoS ONE*, *6*(6), e20387.
- Ginestet, E., Phénix, T., Diard, J., & Valdois, S. (2019). Modeling the length effect for words in lexical decision: The role of visual attention. *Vision Research*, *159*, 10–20.
- Ginestet, E., Valdois, S., & Diard, J. (2022). Probabilistic modeling of orthographic learning based on visual attention dynamics. *Psychonomic Bulletin and Review*, *29*, 1649–1672.
- Gomez, P., Ratcliff, R., & Perea, M. (2008). The Overlap model: A model of letter position coding. *Psychological Review*, *115*(3), 577–601.
- Grainger, J., Dufau, S., & Ziegler, J. C. (2016). A vision of reading. *Trends in Cognitive Sciences*, *20*(3), 171–179.
- Grainger, J., Granier, J.-P., Farioli, F., Van Assche, E., & van Heuven, W. J. B. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 865–84.
- Grainger, J., & Jacobs, A. M. (1994). A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1158–1176.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*(3), 518–565.
- Grainger, J., & van Heuven, W. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *Mental lexicon: “some words to talk about words”* (pp. 1–23). Nova Science Publishers.
- Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 54, 1–13.
- Guerrera, C., & Forster, K. I. (2008). Masked form priming with extreme transposition. *Language and Cognitive Processes*, *23*(1), 117–142.
- Habekost, T. (2015). Clinical TVA-based studies: A general review. *Frontiers in Psychology*, *6*, 290.
- Habekost, T., & Bundesen, C. (2003). Patient assessment based on a theory of visual attention (TVA): Subtle deficits after a right frontal-subcortical lesion. *Neuropsychologia*, *41*, 1171–1188.
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334–337.
- He, Y., & Legge, G. E. (2017). Linking crowding, visual span and reading. *Journal of Vision*, *17*, 1–15.
- Holmes, V., & O’Regan, J. K. (1987). Decomposing French words. In J. K. O’Regan & A. Lévy-Schoen (Eds.), *Eye movements: From physiology to cognition* (pp. 459–466). L. Erlbaum Associates.
- Huckauf, A., Heller, D., & Nazir, T. A. (1999). Lateral masking: Limitations of the feature interaction account. *Perception & Psychophysics*, *61*(1), 177–189.
- Huckauf, A., & Nazir, T. A. (2007). How odgcrnwi becomes crowding: Stimulus-specific learning reduces crowding. *Journal of Vision*, *7*(2), 18, 1–12.
- Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, *22*, 517–560.
- Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1174–1188.
- Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-p: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 147–188). L. Erlbaum Associates.
- Johnston, J. C., & McClelland, J. L. (1974). Perception of letters in words: Seek not and ye shall find. *Science*, *184*(4142), 1192–1194.
- Johnston, J. C. (1978). A test of the Sophisticated Guessing Theory of word perception. *Cognitive Psychology*, *10*(2), 123–153.
- Joo, S. J., White, A. L., Strodtman, D. J., & Yeatman, J. D. (2018). Optimizing text for an individual’s visual system: The contribution of visual crowding to reading difficulties. *Cortex*, *103*, 291–301.
- Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition*, *55*, 332–340.
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304.
- Kinoshita, S., & Norris, D. (2009). Transposed-letter priming of prelexical orthographic representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 1–18.
- Kinoshita, S., & Norris, D. (2012). Task-dependent masked priming effects in visual word recognition. *Frontiers in Psychology*, *3*, 178, 1–12.

- Kinoshita, S., & Norris, D. (2013). Letter order is not coded by open bigrams. *Journal of Memory and Language*, *69*, 135–150.
- Kwon, M., Legge, G. E., & Dubbels, B. R. (2007). Developmental changes in the visual span for reading. *Vision Research*, *47*, 2889–2900.
- LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, *96*(1), 101–124.
- LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323.
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, *111*(4), 880–913.
- Lassus-Sangosse, D., N’guyen-Morel, M.-A., & Valdois, S. (2008). Sequential or simultaneous visual processing deficit in developmental dyslexia? *Vision Research*, *48*, 979–988.
- Lebeltel, O., Bessi re, P., Diard, J., & Mazer, E. (2004). Bayesian robot programming. *Autonomous Robots*, *16*(1), 49–79.
- Lien, M.-C., Ruthruff, E., Kouchi, S., & Lachter, J. (2010). Even frequent and expected words are not identified without spatial attention. *Attention, Perception, & Psychophysics*, *72*(4), 973–988.
- Liu, T., de Schotten, M. T., Alatrelli, I., Ramus, F., & Zhao, J. (2022). Neural dissociation of visual attention span and phonological deficits in developmental dyslexia: A hub-based white matter network analysis. *Human Brain Mapping*, *43*(17), 5210–5219.
- Lobier, M., Peyrin, C., Le Bas, J.-F., & Valdois, S. (2012). Pre-orthographic character string processing and parietal cortex: A role for visual attention in reading? *Neuropsychologia*, *50*, 2195–2204.
- Lobier, M., Peyrin, C., Pichat, C., Le Bas, J.-F., & Valdois, S. (2014). Visual processing of multiple elements in the dyslexic brain: Evidence for a superior parietal dysfunction. *Frontiers in Human Neuroscience*, *8*, 479, 1–16.
- Lupker, S. J., & Davis, C. J. (2009). Sandwich priming: A method for overcoming the limitations of masked priming by reducing lexical competitor effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *35*(3), 618–639.
- Madec, S., Rey, A., Dufau, S., Klein, M., & Grainger, J. (2012). The time course of visual letter perception. *Journal of Cognitive Neuroscience*, *24*(7), 1645–1655.
- Marr, D. (1982). *Vision. a computational investigation into the human representation and processing of visual information*. W.H. Freeman; Company.
- Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. (2009). Crowding, reading, and developmental dyslexia. *Journal of Vision*, *9*(4), 1–18.
- Martens, V. E., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language*, *98*, 140–149.
- Mason, M. (1975). Reading ability and letter search time: Effects of orthographic structure defined by single letter positional frequency. *Journal of Experimental Psychology: General*, *104*(2), 146–166.
- Massaro, D. W. (1973). Perception of letters, words, and nonwords. *Journal of Experimental Psychology*, *100*(2), 349–353.
- McCann, R. S., Folk, C. L., & Johnston, J. C. (1992). The role of spatial attention in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1015–1029.
- McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: A historical and tutorial review. *Frontiers in Psychology*, *4*, 503.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review*, *88*(5), 375–407.
- McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Visual Cognition*, *13*(1), 89–98.
- Montani, V., Facoetti, A., & Zorzi, M. (2015). The effect of decreased interletter spacing on orthographic processing. *Psychonomic Bulletin and Review*, *22*, 824–832.
- Montant, M., Nazir, T. A., & Poncet, M. (1998). Pure alexia and the viewing position effect in printed words. *Cognitive Neuropsychology*, *15*(1-2), 93–140.
- Mozer, M. C., & Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, *2*(2), 96–123.
- Mueller, S. T., & Weidemann, C. T. (2012). Alphabetic letter identification: Effects of perceivability, similarity, and bias. *Acta Psychologica*, *139*, 19–37.
- Murphy, K. (2002, July). *Dynamic Bayesian networks: Representation, inference and learning* [Ph.D. thesis]. University of California, Berkeley.
- Nazir, T. A. (1991). On the role of refixations in letter strings: The influence of oculomotor factors. *Perception & Psychophysics*, *49*(4), 373–389.
- Nazir, T. A., Heller, D., & Sussmann, C. (1992). Letter visibility and word recognition: The optimal viewing



- position in printed words. *Attention, Perception, & Psychophysics*, 52(3), 315–328.
- Nazir, T. A., Jacobs, A. M., & O'Regan, J. K. (1998). Letter legibility and visual word recognition. *Memory & Cognition*, 26(4), 810–821.
- Nazir, T. A., O'Regan, J. K., & Jacobs, A. M. (1991). On words and their letters. *Bulletin of the Psychonomic Society*, 29(2), 171–174.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining the word length effect in visual word recognition: New evidence from the english lexicon project. *Psychonomic Bulletin and Review*, 13(1), 45–52.
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal bayesian decision process. *Psychological Review*, 113(2), 327–357.
- Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, 17(10), 517–524.
- Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General*, 137(3), 434–455.
- Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review*, 119(3), 517–545.
- Norris, D., Kinoshita, S., & Casteren, M. v. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language*, 62, 254–271.
- O'Regan, J. K., & Jacobs, A. M. (1984). Optimal viewing position effect in word recognition: A challenge to current theories. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 185–197.
- O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugailière, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, 10(2), 250–257.
- Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, 89(5), 573–594.
- Pasqualotto, A., Altarelli, I., De Angeli, A., Menestrina, Z., Bavelier, D., & Venuti, P. (2022). Enhancing reading skills through a video game mixing action mechanics and cognitive training. *Nature Human Behaviour*, 6, 545–554.
- Paterson, K., & Jordan, T. (2010). Effects of inter-letter spacing on word identification and eye guidance during reading. *Memory and Cognition*, 38(4), 502–512.
- Patri, J.-F., Diard, J., & Perrier, P. (2015). Optimal speech motor control and token-to-token variability: A Bayesian modeling approach. *Biological Cybernetics*, 109(6), 611–626.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46, 4646–4674.
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4, 1136–1169.
- Pelli, D. G., & Tillman, K. A. (2007). Parts, wholes, and context in reading: A triple dissociation. *PLoS ONE*, 2(8), e680.
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11(10), 1129–1135.
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7(2), 20, 1–36.
- Perea, M., & Gomez, P. (2011). Increasing inter-letter spacing facilitates encoding of words. *Psychonomic Bulletin and Review*, 137(19), 332–338.
- Perea, M., & JLupker, S. (2003). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: State of the art* (pp. 97–120). Psychology Press.
- Perea, M., & Lupker, S. J. (2004). Can *CANISO* activate *CASINO*? transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51, 231–246.
- Perea, M., Moret-Tatay, C., & Gomez, P. (2011). The effects of inter-letter spacing in visual word recognition. *Acta Psychologica*, 137(4), 345–351.
- Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 767–779.
- Peressotti, F., & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception & Psychophysics*, 61(4), 691–706.
- Pernet, C., Valdois, S., Celsis, P., & Démonet, J.-F. (2006). Lateral masking, levels of processing and stimulus category: A comparative study between normal and dyslexic readers. *Neuropsychologia*, 44, 2374–2385.
- Perry, C., & Long, H. (2022). What is going on with visual attention in reading and dyslexia? a critical review of recent studies. *Brain Sciences*, 12, 87.
- Perry, C., & Ziegler, J. C. (2002). Cross-language computational investigation of the length effect in reading

- aloud. *Journal of Experimental Psychology: Human perception and performance*, 28(4), 273–315.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, 114(2), 273–315.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, 61(2), 106–51.
- Perry, C., Ziegler, J. C., & Zorzi, M. (2014). When silent letters say more than a thousand words: An implementation and evaluation of CDP++ in French. *Journal of Memory and Language*, 72, 98–115.
- Peyrin, C., Démonet, J.-F., N’guyen-Morel, M.-A., Le Bas, J.-F., & Valdois, S. (2011). Superior parietal lobule dysfunction in a homogeneous group of dyslexic children with a visual attention span disorder. *Brain & Language*, 118, 128–138.
- Peyrin, C., Lallier, M., Démonet, J.-F., Pernet, C., Baciú, M., Le Bas, J.-F., & Valdois, S. (2012). Neural dissociation of phonological and visual attention span disorders in developmental dyslexia: fMRI evidence from two case reports. *Brain & Language*, 120, 381–394.
- Phénix, T. (2018). *Modélisation bayésienne algorithmique de la reconnaissance visuelle de mots et de l’attention visuelle* [Ph.D. thesis]. Univ. Grenoble Alpes.
- Phénix, T., Diard, J., & Valdois, S. (2016). Les modèles computationnels de lecture. In M. Sato & S. Pinto (Eds.), *Traité de neurolinguistique* (pp. 167–182). De Boeck supérieur.
- Phénix, T., Valdois, S., & Diard, J. (2018). Reconciling opposite neighborhood frequency effects in lexical decision: Evidence from a novel probabilistic model of visual word recognition. In T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp. 2238–2243). Cognitive Science Society.
- Pitchford, N. J., Ledgeway, T., & Masterson, J. (2008). Effect of orthographic processes on letter position encoding. *Journal of Research in Reading*, 31(1), 97–116.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115.
- Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A computational model of the self-teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive Science*, 1–49.
- Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, 7(5), 187–189.
- Rabiner, L. R., & Juang, B.-H. (1993). *Fundamentals of speech recognition*. Prentice Hall.
- Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, 41, 211–236.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81(2), 275–280.
- Reichle, E. D. (2021). *Computational models of reading: A handbook*. Oxford University Press.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, 39, 4403–4411.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26, 445–526.
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using e-z reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16(1), 1–21.
- Reilhac, C., Peyrin, C., Démonet, J.-F., & Valdois, S. (2013). Role of the superior parietal lobules in letter-identity processing within strings: fMRI evidence from skilled and dyslexic readers. *Neuropsychologia*, 51, 601–612.
- Reinhart, S., Schaadt, A. K., Adams, M., Leonhardt, E., & Kerkhoff, G. (2013). The frequency and significance of the word length effect in neglect dyslexia. *Neuropsychologia*, 51, 1273–1278.
- Risko, E. F., Lanthier, S. N., & Besner, D. (2011). Basic processes in reading: The effect of interletter spacing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37(6), 1449–1457.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. the contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89(1), 60–94.
- Russell, S., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Prentice Hall Series in Artificial Intelligence.

- Saghiran, A., Valdois, S., & Diard, J. (2020). Simulating length and frequency effects across multiple tasks with the bayesian braid-phon model. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd annual conference of the cognitive science society*. Cognitive Science Society.
- Scaltritti, M., & Balota, D. A. (2013). Are all letters really processed equally and in parallel? further evidence of a robust first letter advantage. *Acta Psychologica*, *144*, 397–410.
- Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes*, *19*(3), 333–367.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.
- Siakaluk, P. D., Sears, C. R., & Lupker, S. J. (2002). Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 661–681.
- Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-Reader: A model of word recognition and eye movements in text reading. *Psychological Review*.
- Spinelli, D., De Luca, M., Judica, A., & Zoccolotti, P. (2002). Crowding effects on word identification in developmental dyslexia. *Cortex*, *38*(2), 179–200.
- Spoehr, K. T., & Smith, E. E. (1975). The role of orthographic and phonotactic rules in perceiving letter patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *104*(1), 21–34.
- Starrfelt, R., Habekost, T., & Gerlach, C. (2010). Visual processing in pure alexia: A case study. *Cortex*, *46*, 242–255.
- Steinhilber, A., Diard, J., Ginestet, E., & Valdois, S. (2023). Visual attention modulates the transition from fine-grained, serial processing to coarser-grained, more parallel processing: A computational modeling study. *Vision Research*, *207*, 108211, 1–15.
- Steinhilber, A., Valdois, S., & Diard, J. (2022). Bayesian comparators: A probabilistic modeling tool for similarity evaluation between predicted and perceived patterns. *44th Annual Meeting of the Cognitive Science Society*, 2264–2270.
- Stevens, M., & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception and Psychophysics*, *65*(1), 133–151.
- Stolz, J. A., & McCann, R. S. (2000). Visual word recognition: Reattending to the role of spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(4), 1320–1331.
- Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, *5*(11), 1024–1037.
- Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, *122*, 2119–2131.
- Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, *9*(1), 40–50.
- Tydgat, I., & Grainger, J. (2009). Serial position effects in the identification of letters, digits, and symbols. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 480–98.
- Valdois, S. (2022). The visual-attention span deficit in developmental dyslexia: Review of evidence for a visual-attention-based deficit. *Dyslexia*, *28*, 397–415.
- Valdois, S., Bidet-Ildei, C., Lassus-Sangosse, D., Reilhac, C., Nguyen-Morel, M.-A., Guinet, E., & Orliaguet, J.-P. (2011). A visual processing but no phonological disorder in a child with mixed dyslexia. *Cortex*, *46*, 1197–1218.
- Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing: An Interdisciplinary Journal*, *16*, 541–572.
- Valdois, S., Lassus-Sangosse, D., Lallier, M., Moreaud, O., & Pisella, L. (2019). What does bilateral damage of the superior parietal lobes tell us about visual attention disorders in developmental dyslexia? *Neuropsychologia*, *130*, 78–91.
- Valdois, S., Peyrin, C., Lassus-Sangosse, D., Lallier, M., Démonet, J.-F., & Kandel, S. (2014). Dyslexia in a French-Spanish bilingual girl: Behavioural and neural modulations following a visual attention span intervention. *Cortex*, *53*, 120–145.
- Valdois, S., Phénix, T., Fort, M., & Diard, J. (2021). Atypical viewing position effect in developmental dyslexia: A behavioural and modelling investigation. *Cognitive Neuropsychology*, *38*, 319–335.
- Valdois, S., Roulin, J.-L., & Bosse, M.-L. (2019). Visual attention modulates reading acquisition. *Vision Research*, *165*, 152–161.
- van den Boer, M., de Jong, P. F., & Meeteren, M. M. H.-v. (2013). Modeling the length effect: Specifying the relation with visual and phonological correlates of reading. *Scientific Studies of Reading*, *17*(4), 243–256.
- van den Boer, M., Elsje, v. B., & de Jong, P. F. (2015). The specific relation of visual attention span with read-

- ing and spelling in dutch. *Learning and individual differences*, 39, 141–149.
- van der Heijden, A. H. C., Malhas, M. S. M., & van den Roovart, B. P. (1984). An empirical interletter confusion matrix for continuous-line capitals. *Perception & Psychophysics*, 35(1), 85–88.
- Veldre, A., Reichle, E. D., Yu, L., & Andrews, S. (2023). Lexical processing across the visual field. *Journal of Experimental Psychology : Human Perception and Performance*, 49(5), 649–671.
- Veldre, A., Yu, L., Andrews, S., & Reichle, E. D. (2020). Towards a complete model of reading: Simulating lexical decision, word naming and sentence reading with über-reader. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 42.
- Vidyasagar, T. R. (2005). Attentional gating in primary visual cortex: A physiological basis for dyslexia. *Perception*, 34, 903–911.
- Waechter, S., Besner, D., & Stolz, J. (2011). Basic processes in reading: Spatial attention as a necessary preliminary to orthographic and semantic processing. *Visual Cognition*, 2(19), 171–202.
- Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1, 59–85.
- Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, 8(2), 221–43.
- Whitney, C. (2018). When serial letter processing implies a facilitative length effect. *Language, Cognition and Neuroscience*, 33(5), 659–664.
- Whitney, C., Bertrand, D., & Grainger, J. (2012). On coding the position of letters in words: A test of two models. *Experimental Psychology*, 59(2), 109–114.
- Whitney, C., & Cornelissen, P. (2008). SERIOL reading. *Language and Cognitive Processes*, 23(1), 143–164.
- Whitney, C., & Lavidor, M. (2005). Facilitative orthographic neighborhood effects: The SERIOL model account. *Cognitive Psychology*, 51(3), 179–213.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15(4), 160–168.
- Yeshurun, Y., & Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. *Journal of Vision*, 10(10), 16, 1–12.
- Yu, D., Cheung, S., Legge, G., & Chung, S. (2007). Effect of letter spacing on visual span and reading speed. *Journal of Vision*, 7(2), 2, 1–10.
- Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1634), 20120397.
- Zoccolotti, P., Luca, M. D., Judica, A., & Spinelli, D. (2008). Isolating global and specific factors in developmental dyslexia: A study based on the rate and amount model (ram). *Experimental Brain Research*, 186, 551–560.
- Zoccolotti, P., Luca, M. D., Pace, E. D., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language*, 93, 369–373.
- Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., Bravar, L., George, F., Pech-Georgel, C., & Ziegler, J. C. (2012). Extra large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, 109, 11455–11459.
- Zoubinetzky, R., Collet, G., Nguyen-Morel, M.-A., Valdois, S., & Serniclaes, W. (2019). Remediation of allophonic perception and visual attention span in developmental dyslexia: A joint assay. *Frontiers in Psychology*, 10, 1502.

## Appendix A

### Joint probability distribution of the BRAID model

The core definition of the BRAID model is the joint probability distribution

$$JD_{BRAID} = P \left( \begin{array}{cccccccc} W^{0:T} & L^{1:T} & \lambda_{L1:N}^{1:T} & P^{0:T} & A^{1:T} & \mu_A^{1:T} & \sigma_A^{1:T} & C_{A1:N}^{1:T} & \lambda_{P1:N}^{1:T} \\ G^{1:T} & S^{1:T} & \Delta I_{1:N}^{1:T} & I_{1:N}^{1:T} & \lambda_{D1:N}^{1:T} & D^{0:T} & C_{D1:N}^{1:T} & & \end{array} \right),$$

which is defined by the following decomposition:

$$JD_{BRAID} = P(W^0)P(D^0) \prod_{n=1}^N P(P_n^0) \prod_{t=1}^T \left[ \begin{array}{l} P(W^t | W^{t-1}) \prod_{n=1}^N P(L_n^t | W^t) \\ P(D^t | D^{t-1}) P(C_{D1:N}^t | D^t) \\ \prod_{n=1}^N \left[ P(\lambda_{Dn}^t | \lambda_{Ln}^t C_{Dn}^t) P(\lambda_{Ln}^t | L_n^t P_n^t) \right] \\ \prod_{n=1}^N P(P_n^t | P_n^{t-1}) \\ P(A^t | \mu_A^t \sigma_A^t) P(\mu_A^t) P(\sigma_A^t) \prod_{n=1}^N P(C_{An}^t | A^t) \\ \prod_{n=1}^N P(\lambda_{Pn}^t | P_n^t I_n^t C_{An}^t) \\ P(G^t) \prod_{n=1}^N \left[ P(S_n^t) P(\Delta I_n^t) P(I_n^t | S_{1:N}^t \Delta I_n^t G^t) \right] \end{array} \right]. \quad (A1)$$

The top terms (first line of Equation (A1), outside of the main product) concern the initial state of the model (at time  $t = 0$ ), whereas the innermost product contains the temporally local portion of the model, i.e., the model that is iterated at each time step  $t \neq 0$ . The innermost product is laid out over seven lines, roughly following a top-down traversal of the dependency structure shown Figure 2, from lexical knowledge to stimulus. The full definition of all terms of Equation (A1) are found elsewhere (Phénix, 2018), and the more relevant terms are explained in the main text.

## Appendix B

### Simulation of benchmark effects

In this section, we present the results of a series of simulations that were carried out to evaluate the capacity of BRAID to account for benchmark effects, namely context effects on letter identification, including the word and pseudoword superiority effect, and their modulation as a function of visual attention, and for transposed-letter priming effects in lexical decision.

## Context effects on letter identification

Since the seminal studies of Reicher (1969) and Wheeler (1970), the two-alternative forced-choice (2AFC) paradigm is used to explore context effects on letter perception. In this paradigm, a stimulus is briefly presented, followed by a mask then a letter-pair located above and below one of the input string position. The participant has to decide which of the two letters was present in the string at this position. For example, the word “LAST” is briefly presented, followed by the O/A letter pair in second position, so that each letter forms an existing word with the surrounding context (“LAST”-“LOST”).

The use of this paradigm revealed two robust phenomena: the word superiority effect – letters are more accurately identified in briefly presented words than in the context of unrelated letter strings (i.e., nonwords) or in isolation (Cattell, 1886; Spoehr & Smith, 1975) –, and the pseudoword superiority effect – letters are identified more accurately when embedded in pseudowords (orthographically legal but meaningless letter strings) than in nonwords (Baron & Thurston, 1973; Carr et al., 1978; Grainger & Jacobs, 1994). The word and pseudoword superiority effects have been explored within the IA framework (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), the AV model (Paap et al., 1982), and DROM (Grainger & Jacobs, 1994). It was further shown that these effects are particular cases of a more general context familiarity effect: a target letter is better recognized when embedded within a letter context that occurs in existing words (Rumelhart & McClelland, 1982).

We perform three simulations, inspired from classical experiments, to show that BRAID accounts for the word superiority effect (Johnston, 1978, Experiment 2) in Simulation 1, for the pseudoword superiority effect (Chase & Tallal, 1990) in Simulation 2, and for more general context familiarity effect (Rumelhart & McClelland, 1982, Experiment 10) in Simulation 3.

**Method.** All three simulations reproduce the conditions of the 2AFC paradigm. Simulations 1 and 2 use the original sets of words. For Simulation 3, as words were not listed in the experiment reported in the original article by Rumelhart and McClelland (1982), the four sets of word pairs from (Johnston, 1978, Experiment 2) were used and pseudowords were generated as described in the original article.

Simulation 1 explores letter identification in the context of words or nonwords, or in isolation (e.g., letter A in “LAST”, “OASW”, or “\_A\_”). Simulation 2 explores letter identification in the context of words (“DARK”), pseudowords (“DARL”), or nonwords (“KADM”). In Simulation 3, letter identification is assessed in the context of pseudowords, word-like nonwords, and QXJ nonwords. Word-like consonant nonwords were generated from words by changing the vowel for a consonant (e.g., “LCST” from

“LAST”) so that the target letter in the nonword (“C” in “LCST”) shared the same consonant context as the target letter in the word (“A” in “LAST”). QXJ nonwords were generated for each target letter by using the letters “Q”, “X”, “J” in various orders as context (e.g., “XCQJ”).

For each simulation, BRAID provides information on the mean time course of letter identification for each context condition, using the inference for letter identification with lexical influence (Equation (4)). The time course from one context condition is then used to identify the iteration at which simulated and experimental data match. Note that the alignment here matches predicted performance, and not response times. Nevertheless, this condition then serves as calibration; the results for the two other conditions at this iteration are therefore model predictions that can be compared with experimental data.

**Simulation results.** Results from Simulation 1 show that BRAID successfully simulates the word superiority effect (see Figure B1 (a)). A repeated measures ANOVA carried out on the simulated data, with context as a within-items variable, shows a significant main effect of context [ $F(2, 542) = 913.7, p < .001, \eta^2 = 0.771$ ]. Target letters are identified more efficiently in the context of words than when presented in isolation [ $t(542) = 28.73, p < .001$ ] and more efficiently when presented in isolation than within nonwords [ $t(542) = -12.71, p < .001$ ]. Exploration of the temporal dynamics of perceptual accumulation further shows that the word superiority effect is quite robust over time. Indeed, the ordering of performance between conditions holds, independently of the chosen reference iteration.

Results from Simulation 2 show that BRAID also simulates the pseudoword superiority effect (see Figure B1 (b)). There is a main effect of context [ $F(2, 158) = 83.65, p < .001, \eta^2 = 0.514$ ], in particular characterized by the fact that letters are more accurately identified within the context of pseudowords than within the context of nonwords [ $t(158) = 4.373, p < .001$ ]. However, BRAID predicts better letter identification within words than within pseudowords, contrary to the human data, in which both are equal. The pseudoword superiority effect is observed all along processing time course in the simulation, and is thus very robust over time.

As shown in Figure B1 (c), in Simulation 3, BRAID well captures the experimental data showing a context familiarity effect. Simulation results show a significant context effect [ $F(2, 158) = 71.55, p < .001, \eta^2 = 0.475$ ]. The target letter is better identified when embedded in word-like consonant than in QXJ nonwords [ $t(158) = 10.06, p < .001$ ]. As in the behavioral experiment, the probability to identify the target letter accurately is almost identical in the context of pseudowords as in the context of word-like consonant nonwords [ $t(158) = -0.47, p = 1$ ]. Exploration of the time course of processing shows that the same pattern

is observed all along processing. The context effect is very stable over time, thus quite robust and largely independent of the selected threshold. Overall, BRAID successfully reproduces the word and pseudoword superiority effect as particular cases of context effects on letter perceptibility.

#### Attention modulation of the word superiority effect: cue condition

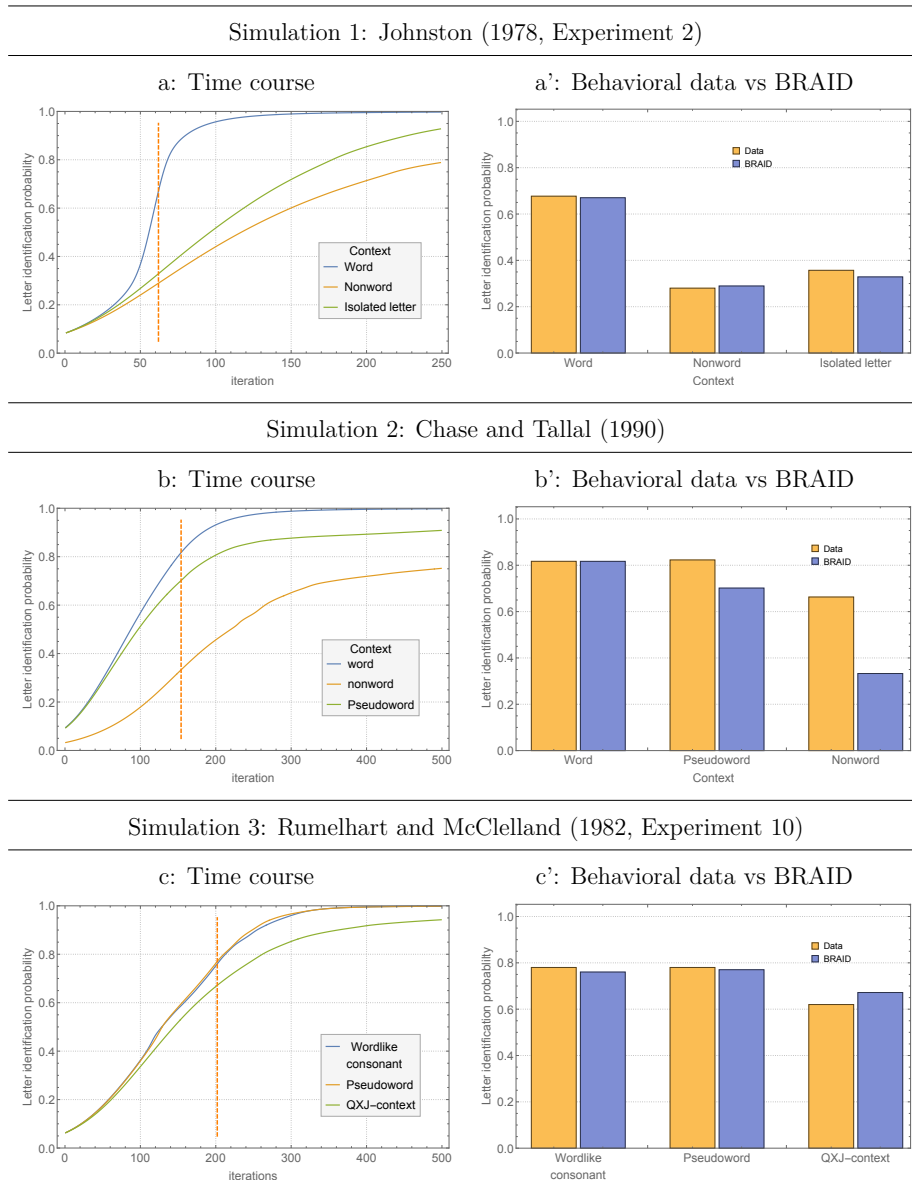
**Method.** In Simulation 4, target letters are presented either in the context of real words (COIN-JOIN) or in the context of non-words (CPRD-JPRD). In the first condition of the behavioral 2AFC task, stimuli were presented centered on the fixation point and participants were instructed to process the whole input letter-string. In the second condition, a cue indicating the position of the target letter was displayed prior to stimulus presentation and the participants were asked to focus attention on the cued position only. For words, the target letter was more accurately identified in the whole string condition than in the cued condition but the reverse was found for non-words. In other words, taken together, these experiments jointly show that the word superiority effect was abolished in the cued condition.

The words and non-words used in Simulation 4 are taken from Simulation 1. The “whole string” condition is simulated using the default parameters of BRAID and the method described in Simulation 1. The “cued condition” is simulated using the same default parameters, except for gaze position  $g^t$  and the mean  $\mu_A^t$  of the attention distribution, which are both aligned with the target letter position. Results of the simulations for the word and non-word stimuli are presented on Figure B2 (a, a', b, and b').

**Simulation results.** BRAID well captures the opposite direction of performance increase for words and non-words: our simulations yield a whole-string condition facilitation for words [ $t(271) = 14.93, p < .001$ ] and a cued-condition facilitation for non-words [ $t(271) = -15.99, p < .001$ ]. In our simulations, as in the behavioral data, the word superiority effect is abolished in the cued condition. Exploration of processing time course shows a very stable advantage of the cued condition over time for the non-words. For words, however, we observe a reversal of the predicted facilitation: at early stages of processing, the cued condition shows faster accumulation of perceptual information, whereas, after iteration 54, the whole-word condition catches up and overpasses the cued condition. The time processing curves show how the word context gradually contributes to letter identification, so much so that the whole word condition becomes facilitating.

#### Attention modulation of the word superiority effect: inter letter spacing and target position predictability

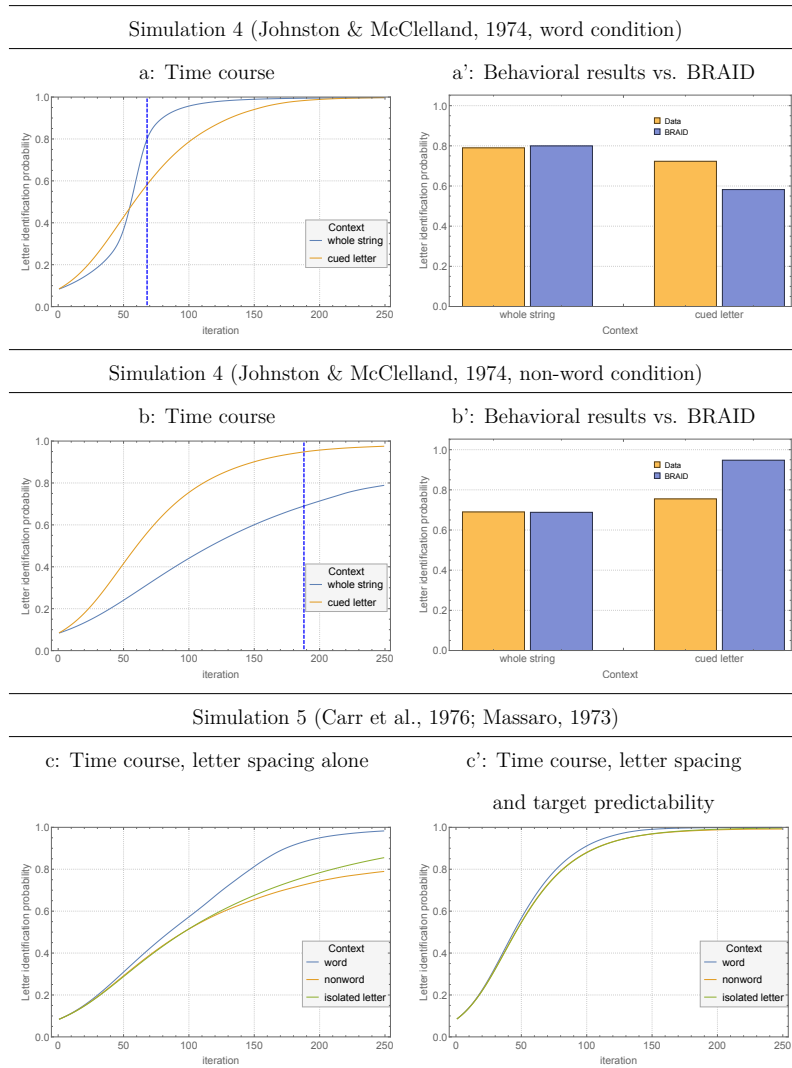
**Method.** In a series of two experiments, Massaro (1973) and Carr et al. (1976) explored letter identification in

**Figure B1**

*Results of Simulations 1, 2, and 3 of the word (top panels) and pseudoword (middle panels) superiority effect, and their generalization to the context familiarity effect (bottom panels). Left panels: time course of probability values for correct letter identification. Vertical dashed orange lines indicate the reference iteration obtained by matching simulated and human data. Right panels: comparison of simulated RTs (blue bars) and behavioral RTs (orange bars) for all experimental conditions; the condition used for selecting the reference iteration is the leftmost, the other two conditions are predictions of the model.*

the context of words (AGE) and non-words (SGL), or when presented in isolation (\_G\_). They used a 2AFC paradigm, varying between-letter spacing and the position of the target letter. In the first experiment (Carr et al., 1976), spacing was standard and the target could randomly occur at each of the three serial positions (P in POT, APE, TAP): this replicated the classical word superiority effect. In the second experi-

ment however (Massaro, 1973), the stimuli were presented with an enlarged letter spacing, while the target letter always occurred in central position: the word superiority effect is abolished in these conditions. Comparison of the two experiments therefore suggests that the word superiority effect may be sensitive to letter spacing and/or target location predictability.

**Figure B2**

Results of Simulations 4 and 5 on modulation of the word superiority effect by experimental conditions. Panels a, b, c, and c' show the time course of probability values for correct letter identification. Vertical dashed blue lines indicate the reference iteration obtained by matching simulated and human data. Panels a' and b': comparison of simulated and human data; the condition used for selecting the reference iteration is to the left, the other conditions are predictions of the model.

In Simulation 5, we assess whether the word superiority effect is abolished by manipulating either letter spacing alone or letter spacing and target predictability simultaneously. The set of words and non-words are taken from Simulation 1. Two parameters were modified to simulate larger spacing between letters: first, the acuity slope parameter ( $\theta_G$ ) was doubled to account for larger eccentricity; second, lateral interference ( $\theta_l$ ) was reduced from 16.25% to 2.5% to simulate reduction of crowding effects when letters are more spaced. To account for target location predictability, the position of gaze ( $g^t$ ) and the mean of the attention distribution ( $\mu_A^t$ ) were aligned on the target letter position, as previously

in Simulation 4.

**Simulation results.** Results of Simulation 5 are presented on Figure B2 (c and c'). For the “large space” condition, time course of processing shows a strong word recognition superiority effect that is stable over time. Letters are identified more accurately in the context of words than in the context of non-words or when presented in isolation. Therefore, the word superiority effect is not sensitive to manipulation of spacing alone. However, when letter spacing is enlarged and when attention focuses on the target letter location, the word superiority effect is abolished. Therefore, our simulations show that the classical word superiority effect



can be modulated by the distribution of visual attention.

### Positional effects in primed lexical decision

A wealth of behavioral experiments has explored transposed-letter priming effects in lexical decision, showing shorter RTs when the target word is primed by a letter string that contains the same letters in a different position (Adelman et al., 2014; Perea & Lupker, 2004; Peressotti & Grainger, 1999; Schoonbaert & Grainger, 2004). Transposed-letter primes produce a strong priming effect, close to that of identity primes (Forster et al., 1987; Perea & JLupker, 2003; Schoonbaert & Grainger, 2004). Transposed letter priming is stronger than substituted letter priming. For example, “JUDGE” (positions noted 12345) is recognized faster when primed by “jugde (positions noted 12435, indicating transposition of positions 3 and 4; here and in the following, we note primes in lower case and targets in upper case) than when primed by “junpe” (12DD5, the “D” indicating substitution by a different letter at that position).

Priming effects have been reported on words of different length, in tasks using primes that differ in the number of transposed letters and in their relative position. Transposed letter priming effects are critical to test the validity of letter position coding assumptions in computational models of word recognition. Slot-based coding models, like the IA model (McClelland & Rumelhart, 1981) or the original version of BR (Norris, 2006) cannot account for such effects, whereas they are well captured by the open-bigram model (Schoonbaert & Grainger, 2004), SERIOL (C. Whitney & Cornelissen, 2008), SCM (Davis, 2010), OVERLAP (Gomez et al., 2008), or the noisy-position version of the BR (Norris et al., 2010).

**Method.** We assess BRAID’s ability to generate a large range of priming effects in lexical decision. Simulation 6 compares conditions of identity, transposed, and substituted letter priming from an experiment by Norris et al. (2010, Experiment 1). Simulation 7 uses the different conditions of extreme transposed-letter priming from an experiment by Guerrero and Forster (2008, Experiments 1A, 2, and 3). We use the same stimuli as in the experiments.

In Simulation 6, the targets are 5-letter English words. Priming effects are studied in conditions of identical (12345, e.g., “under”-“UNDER”; “ID” condition) and adjacent transposed-letter primes (13245 or 12435, e.g., “udner”-“UNDER”; “TL” condition), in conditions of substituted-letter priming by either letter replacement (1DD45 or 12DD5, e.g., “ulger”-“UNDER”; “SL” condition) or letter repetition (11445 or 12255, e.g., “uueer”-“UNDER”; “repSL” condition), and a control condition of unrelated primes (DDDDD, e.g., “gypny”-“UNDER”; “ALD” (for “all-different”) condition).

In Simulation 7, the targets are 8-letter words and conditions of extreme transpositions are used while

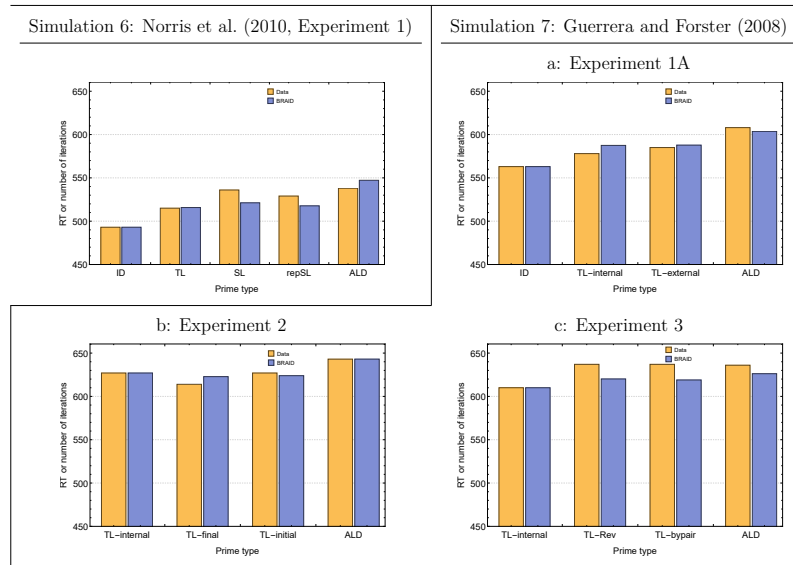
varying the relative position of the transposed letters. In simulation 7a (Guerrera & Forster, 2008, Experiment 1A), the transpositions are either internal (13254768, e.g., “anbroaml”-“ABNORMAL”; “TL-internal” condition) or external (21345687, e.g., “banormla”-“ABNORMAL”; “TL-external” condition). In simulation 7b (Guerrera & Forster, 2008, Experiment 2), letter pairs are reversed while preserving the first two (12436587, e.g., “abonmrla”-“ABNORMAL”; “TL-final” condition) or the last two letters (21436578, “baonmral”-“ABNORMAL”; “TL-initial” condition) in the final and initial transposed-letter priming conditions. In simulation 7c (Guerrera & Forster, 2008, Experiment 3), conditions of reverse halves (43218765, e.g., “onbalarm”-“ABNORMAL”; “TL-Rev” condition) or total by-pair transposition (21436587, e.g., “baonmrla”-“ABNORMAL”; “TL-bypair” condition) are further assessed. An ID condition with an identity prime is further used in Simulation 7a, and a TL-internal condition in which all six internal letters are transposed (e.g., “anbroaml”-“ABNORMAL”) is used in Simulations 7b and 7c. Finally, Simulations 7a, 7b, and 7c also feature an ALD, control condition (DDDDDDDD, e.g., “whiscoon”-“ABNORMAL”).

Presentation duration of the prime is similar in the model as in the experiments (53 and 40 iterations for Simulation 6 and 7 respectively). The threshold value was set to 0.72 for Simulation 6 and to 0.94 for all three sets of conditions of Simulation 7.

**Simulation results.** All target words were accurately recognized in simulation 6 and more than 90% in the different conditions of Simulation 7 (94.8% in Simulation 7a and 7c; 98.9% in Simulation 7b). Results of the priming effects on RTs in lexical decision are presented in Figure B3, for the simulations and the experiments.

Results of Simulation 6 provide a good fit to the data ( $R^2 = .75$ ). The main effect of prime type is significant [ $F(4, 396) = 2165, p < .001, \eta^2 = 0.956$ ]. BRAID reproduces priming effects of remarkably similar direction and amplitude as in the experiment: larger priming effect of the identity than the unrelated prime (–54 iterations vs. –45 ms, [ $t(99) = 60.28, p < .001$ ]), of the transposed-letter prime than the unrelated prime (–32 iterations vs. –23 ms, [ $t(99) = 54.76, p < .001$ ]) and larger priming for the identity than the transposed-letter condition (–23 iterations vs. –22 ms, [ $t(99) = 40.04, p < .001$ ]). As in the experiment, the substitution and letter-repetition priming conditions have similar effects on word recognition. However, the model predicts significant priming effects in these two conditions of substitution and letter repetition compared to the unrelated prime (–26 iterations and –30 iterations, all  $ps < .001$ ), which are not observed in the experiment (–2 ms and –9 ms).

Results of Simulation 7 (Figure B3) show quite a good fit of the data for both Experiment 1A (Simulation 7a:  $R^2 = .91$ ) and Experiment 2 (Simulation 7b:  $R^2 = .83$ ), and a



**Figure B3**

Results of Simulations 6 and 7 of the positional effects in primed lexical decision. Each panel compares simulated RTs (blue bars) and behavioral RTs (orange bars). Simulated RTs are aligned with behavioral RTs on the leftmost condition in each experiment. See text for condition nomenclature.

slightly lower fit for Experiment 3 (Simulation 7c:  $R^2 = .75$ ). As in Experiment 1A, a reliable internal (e.g., “anbroaml”-“ABNORMAL”; priming =  $-16$  iterations vs.  $-30$  ms,  $[t(85) = 28.25, p < .001]$ ) and external (e.g., “banormla”-“ABNORMAL”; priming =  $-16$  iterations vs.  $-23$  ms,  $[t(85) = 28.60, p < .001]$ ) transposed-letter priming effect is simulated. The identity prime produces the larger priming effect in Simulation 7a as in the experiment ( $-41$  iterations vs.  $-45$  ms,  $[t(85) = 50.41, p < .001]$ ). The pattern of simulated results in Experiment 2 is quite similar to that of the behavioral data. The model produces reliable priming effects in all three conditions of transposition, that is to say, when only the two first (e.g., “abonmrla”-“ABNORMAL”; priming =  $-20$  iterations vs.  $-29$  ms), or the two last letters are preserved (e.g., “baonmral”-“ABNORMAL”; priming =  $-19$  iterations vs.  $-16$  ms), or when only the first and last letters are preserved (e.g., “anbroaml”-“ABNORMAL”; priming =  $-16$  iterations vs.  $-16$  ms), all  $ps < .001$ . Simulation of the extreme transposition conditions of Experiment 3 shows only minimal priming effects of a few milliseconds and iterations in conditions of letter reversed halves (e.g., “onbalamr”-“ABNORMAL”, priming =  $-6$  iterations vs.  $+1$  ms) and total by-pair letter transposition

(e.g., “baonmrla”-“ABNORMAL”, priming =  $-7$  iterations vs.  $+1$  ms). Contrary to the behavioral data, these effects are significant ( $[t(85) = 12.5, p < .001]$  and  $[t(85) = 17.53, p < .001]$  respectively). Finally, there is a reliable priming effect in the condition in which all six internal letters are transposed (e.g., “anbroaml”-“ABNORMAL”; priming =  $-16$  iterations vs.  $-26$  ms,  $[t(85) = 28.25, p < .001]$ ).

Overall, the model accounts for a broad range of transposed-letter priming effects, with few discrepancies between simulations and empirical data. However, we also note some discrepancy between experimental observations themselves. For instance, BRAID predicts significant priming effects when using primes with 2-letter substitutions, against the experimental findings reported by Norris et al. (2010, Experiment 1), but the simulated results are in line with the significant priming effects reported by Adelman et al. (2014) in similar conditions but for 6-letter words. Furthermore, while Guerrero and Forster (2008) report a nonsignificant negative effect of priming (of  $+1$  ms) in the condition of total by-pair letter transposition, a priming effect of a size ( $-9$  ms) similar to the one simulated by BRAID was reported by Lupker and Davis (2009) in a replication of the initial experiment.