COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12917

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12917

Friends in Low-Entropy Places: Orthographic Neighbor Effects on Visual Word Identification Differ Across Letter Positions

Sahil Luthra,[a,b] Heejo You,[a] Jay G. Rueckl,[a,b,c] James S. Magnuson[a,b,c]

[a]*Department of Psychological Sciences, University of Connecticut*
[b]*Connecticut Institute for the Brain and Cognitive Sciences*
[c]*Haskins Laboratories*

Received 16 December 2019; received in revised form 24 June 2020; accepted 1 October 2020

Abstract

Visual word recognition is facilitated by the presence of *orthographic neighbors* that mismatch the target word by a single letter substitution. However, researchers typically do not consider *where* neighbors mismatch the target. In light of evidence that some letter positions are more informative than others, we investigate whether the influence of orthographic neighbors differs across letter positions. To do so, we quantify the number of *enemies* at each letter position (how many neighbors mismatch the target word at that position). Analyses of reaction time data from a visual word naming task indicate that the influence of enemies differs across letter positions, with the negative impacts of enemies being most pronounced at letter positions where readers have low prior uncertainty about which letters they will encounter (i.e., positions with low entropy). To understand the computational mechanisms that give rise to such positional entropy effects, we introduce a new computational model, VOISeR (Visual Orthographic Input Serial Reader), which receives orthographic inputs in parallel and produces an over-time sequence of phonemes as output. VOISeR produces a similar pattern of results as in the human data, suggesting that positional entropy effects may emerge even when letters are not sampled serially. Finally, we demonstrate that these effects also emerge in human subjects' data from a lexical decision task, illustrating the generalizability of positional entropy effects across visual word recognition paradigms. Taken together, such work suggests that research into orthographic neighbor effects in visual word recognition should also consider differences between letter positions.

Keywords: Visual word recognition; Orthographic neighbor; Letter position; Entropy; Computational modeling

Correspondence should be sent to Sahil Luthra, Department of Psychological Sciences, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020. E-mail: sahil.luthra@uconn.edu

1. Introduction

A hallmark of visual word identification is that presentation of a target word entails not only accessing the target but also activating orthographically related words. A word's *orthographic neighbors* (henceforth, simply *neighbors*) are typically defined as words that differ from the target word by the substitution of a single letter (Coltheart, Davelaar, Jonasson, & Besner, 1977). For instance, the neighborhood for *WORD* includes *CORD*, *WARD*, *WOOD*, and *WORM*. In general, words from larger neighborhoods are recognized more quickly than words with fewer neighbors (Andrews, 1997), though the precise way in which neighbors influence the dynamics of lexical access is fairly complex.

To understand how neighborhood effects unfold, consider the interactive activation model of word recognition proposed by McClelland and Rumelhart (1981). This model includes bottom-up connections from letter units to word units, reciprocal top-down connections from word units to letter units, and inhibitory lateral connections at both the letter level and word level; notably, all letter nodes are activated in parallel, rather than in a strictly serial (e.g., left-to-right) fashion. When the model encounters a target word (such as *WORD*), neighboring words (e.g., *WORM*) are partially activated. The presence of top-down connections in the model's architecture means that every neighbor reinforces its constituent letters, and because most of these letters match the target (*W*, *O*, and *R*), they "confirm" the lexical prediction by subsequently boosting activation of the target through bottom-up connections. At the same time, every neighbor (*WORM*, *CORD*, etc.) directly inhibits the target (*WORD*) through inhibitory lateral connections at the word level. In this way, neighbors can both inhibit and/or enhance the target word, even if the net effect of having multiple neighbors tends to be facilitative.

In assessing the influence of neighbors on word recognition, it may be useful to consider not only the size of the neighborhood but also its composition. To this end, every neighbor can be classified as a *friend* or *enemy* of a given letter position, depending on whether it respectively matches or mismatches the target word at that position.[1] The word *FORD*, for instance, is an enemy of *WORD* at the first letter position but a friend at positions 2, 3, and 4. Note that each neighbor is a friend at more positions than it is an enemy (i.e., every neighbor is a friend at all but one position, and each neighbor is an enemy at only one position). It is through this "relative friendliness" that orthographic neighbors facilitate recognition of a target word.

A consideration of letter position allows for a more nuanced characterization of the dynamics of visual word recognition. One way to consider letter position is through a metric termed neighborhood *spread*, defined as the number of letter positions where substitutions can yield neighbors (Johnson & Pugh, 1994; Pugh, Rexer, Peter, & Katz, 1994). Critically, this metric is independent from the total number of neighbors. For instance, *WORD* has a spread of four (as shown in Fig. 1), but *BOOK* has a spread of three (as no substitution at the second position can yield a neighbor). The importance of spread was demonstrated convincingly in a study by Mathey and Zagar (2000), who found that readers were slower to recognize words with a low spread of neighbors compared to those with a higher spread. Such an effect can be understood by appealing to the concept of "gang effects," as described by McClelland and Rumelhart (1981). When a set of neighbors all mismatch a target word at the same letter position, these enemies form a "gang" of words that mutually reinforce each other through their interactions with letter nodes; for instance, all the words that mismatch the target *WORD* in word-initial position constitute an *_ORD* gang. Because of mutual reinforcement among gang members, words in large gangs tend to be relatively more activated than words in small gangs. If a target word has low spread, its neighbors will be strongly activated and will strongly inhibit the recognition of the target. By comparison, if a target word has the same number of neighbors but the neighbors are spread out over relatively many letter positions, the neighbors will not be as highly activated, and the target word will thus be recognized relatively quickly. More generally, such findings suggest that the relative number of friends and enemies at each position has important consequences for the activation dynamics in word identification. An unanswered question is whether it matters *where* these friends and enemies reside. In other words, does the facilitative influence of orthographic neighbors depend on *which* specific positions where substitutions can yield neighbors?

The idea that it may be important to consider the relative location of friends and enemies has its roots in visual word identification studies, suggesting that some letter positions are relatively more important than others. A number of studies have suggested that

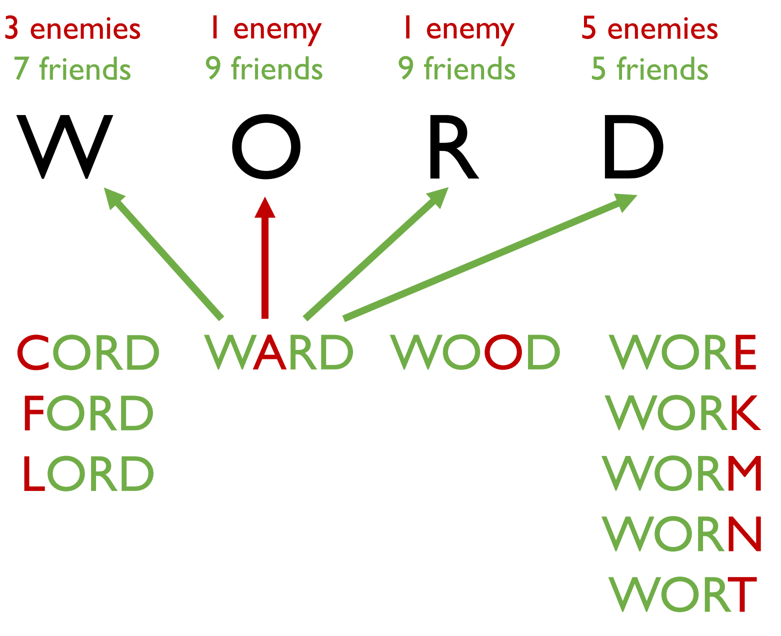

Fig. 1. The word *WORD* has 10 (substitution) orthographic neighbors, and every neighbor can be classified as a friend or as an enemy at each letter position. For instance, the neighbor *WARD* is an enemy to *WORD* at position 2 (red arrow) but a friend at positions 1, 3, and 4 (green arrows). By definition, the number of friends at a given position is equal to the total number of enemies at all other positions. For instance, the number of friends at the second position (9 friends) is equal to the total number of enemies at the other positions (3 + 1 + 5).

access to the initial letter positions of a word is particularly important; for instance, a study by Inhoff and Tousman (1990) found that across both word naming and lexical decision tasks, there were stronger benefits of primes that share word-initial letters (e.g., *BITXXX*) than primes that share word-final letters (e.g., *XXXTER*) on later recognition of visual target words (e.g., *BITTER*). Biases for word-initial positions appear to reflect the fact that initial positions often have more informative (i.e., less predictable) letters rather than reflecting a pre-lexical strategy of always scanning from left to right (Grainger & Jacobs, 1993; O'Regan, Lévy-Schoen, Pynte, & Brugaillière, 1984). For instance, O'Regan et al. (1984) found that when given words where the more informative letters were near the beginning of the word, participants read the words more quickly if initial fixations were directed to early letters as compared to later ones. However, the optimality of the early viewing position was reduced when readers were given words where the more informative letters were near the end of the word. Furthermore, when asked to reread the words with word-final informative letters, reading times were slightly faster when participants initially fixated on the end of the word as compared to the beginning. Taken together, such results suggest that the way in which letter position effects emerge in visual word recognition may be fundamentally tied to how much information is provided at each position.

Because visual word identification is facilitated when readers initially fixate on relatively more informative letter positions, Blais et al. (2009) suggested that an optimal reading strategy might be to process letter positions in accordance with their information content. To determine the letter positions that should be prioritized, the authors implemented an "ideal reader" model that sampled letters serially, letter by letter, but instead of sampling them in a simple left-to-right manner, as other models of reading have assumed (Coltheart & Rastle, 1994; Whitney, 2008), the model iteratively identified which letter position to process next based on which position would most reduce remaining uncertainty about word identity. By identifying the optimal reading strategy for a large set of French and English words, the authors computed a *relative importance* metric for each letter position in four-, five-, six-, and seven-letter words. Conceptually, this metric reflects the importance of accessing a particular letter position early in processing. To test whether human behavior approximated the optimal reading strategy, Blais et al. had a group of participants complete a speeded naming task with five-letter French words. Critically, on each trial, a movie of semi-transparent "bubbles" was overlaid on the word, such that different letter positions were briefly obscured at different points in time. By sampling across a range of trials, the authors were able to ascertain that efficient reading of five-letter words depended on early access to positions 1, 3, and 4—the positions that their ideal reader analysis had identified as the most important for words of this length. From this, the authors concluded that readers process letters in accordance with their relative importance rather than processing all letters in parallel; Blais et al. posited that readers either process letters serially (in order of how much they reduce uncertainty about word identity) or that they process letters in a partly parallel fashion that still considers more informative letter positions prior to less informative ones.

In the current study, we investigated how orthographic neighborhood effects in visual word recognition might differ based on neighborhood composition. Rather than considering the overall number of neighbors a word might have (i.e., how many words mismatch the target at some position), we considered how many neighbors are enemies of the target at each letter position (i.e., how many words mismatch the target at that particular position). In this way, we were able to assess whether neighborhood effects differ in strength between letter positions. Finally, we considered whether positional differences in neighborhood effects were related to the amount of information generally provided by knowing the letter identity at each position (i.e., how much uncertainty there is about a letter's identity).

This paper is structured as follows: In Section 2, we present analyses of archival data from a speeded word naming task, collected as part of the multi-university English Lexicon Project (ELP; Balota et al., 2007). In Section 3, we present results from a novel computational model of word naming, VOISeR (Visual Orthographic Input Serial Reader); to the extent that the model's performance resembles human performance, VOISeR can provide insight into the computations that might underlie any effects observed in Section 2. Following this, we consider data from a lexical decision task that was also collected through the ELP (Section 4). We close with a general discussion of all our results (Section 5).

2. Word naming (human subjects)

We first consider how neighborhood effects might differ as a function of letter position in a speeded word naming task. Data were obtained from the ELP database, a full overview of which is provided by Balota et al. (2007). In brief, the database includes lexical characteristics (word frequency, neighborhood size, etc.) for 39,313 words as well as trial-by-trial data from two visual word recognition tasks (speeded word naming and lexical decision).

2.1. Methods

Analyses were limited to words that were three to eight letters in length, resulting in 23,709 total words. Altogether 471 participants contributed to the dataset analyzed here. We only analyzed data from trials in which the participant self-reported that they had identified the word correctly, yielding an average of 27.3 ($SD = 3.5$) observations per item.

For each word, we computed the number of enemies at each letter position (i.e., how many substitution neighbors mismatched the target word at that position). In Appendix A.1, we describe how to restate our results in terms of the number of friends.

We also quantified the amount of information provided by each position by determining how much a priori uncertainty there is about the letter identity at each position. To do so, we computed the (*Shannon*) *entropy* at each letter position based on all the words of that length in the ELP (Fig. 2). Entropy is a metric used in information theory to establish how much information is provided by an event. It is calculated as follows (where x_i represents each possible letter at a given position):

$$\text{Entropy} = -\sum p(x_i) \times \log_2 p(x_i).$$

The smaller the value, the less uncertainty there is about the letter's identity (again, given only information about word length). If only one letter were possible at a given position (e.g., if every word in the English language began with an *e*), then the entropy at that position would be 0. Similarly, if all letters were equally probable at a given position, the entropy would be ∞. In some letter positions, such as the first position in a three-letter word (Fig. 2A), probability is relatively well distributed among the possible letter identities. At these positions, there is relatively high uncertainty (high entropy) about the identity of this letter, and learning the letter identity at this position is highly informative with regard to word identity. By contrast, there are other positions, such as the second position in a three-letter word (Fig. 2B), where probability tends to be amassed on just a few possible letter identities. For such low-entropy positions, there is relatively little uncertainty about the identity of the letter, and learning the letter identity is not very informative. In Fig. 2C, we provide the entropy values for each of position in three-, four-, five-, six-, seven-, and eight-letter words. (A dark square indicates that a relatively large amount of probability is amassed on that particular letter identity. To link panels A and B to panel C, first note that probability is indicated by bar height in panels A and B, but by grayscale [darker = higher probability] in panel C. Thus, panel A corresponds to the top row of the top left subpanel of panel C, and panel B corresponds to the row below in the same subpanel.)

Note that our entropy measure is conditioned only on the length of the word; that is, it reflects the a priori amount of uncertainty about letter identity given only the length of the word. This is in contrast to the *relative importance* metric used by Blais et al. (2009); while that study did use a similar calculation to quantify uncertainty, the authors conditioned their calculations not only on word length but also on the identities of letters that had already been processed in the word. Such an approach assumes a serial processing system (Coltheart & Rastle, 1994; Whitney, 2008), though many models of visual word recognition instead assume that inputs are processed in parallel (McClelland & Rumelhart, 1981). In choosing to measure a priori uncertainty, we do not make strong commitments to a model that samples input serially.[2]

2.2. Results

We conducted a series of regression analyses to assess how the number of enemies at each position predicted response times in the naming task; a separate regressor was used for each position, and a natural log transformation was applied to response time data to improve fit to a normal distribution. Note that because position-specific statistics are contingent on word length, separate models were used for words of different length. Only

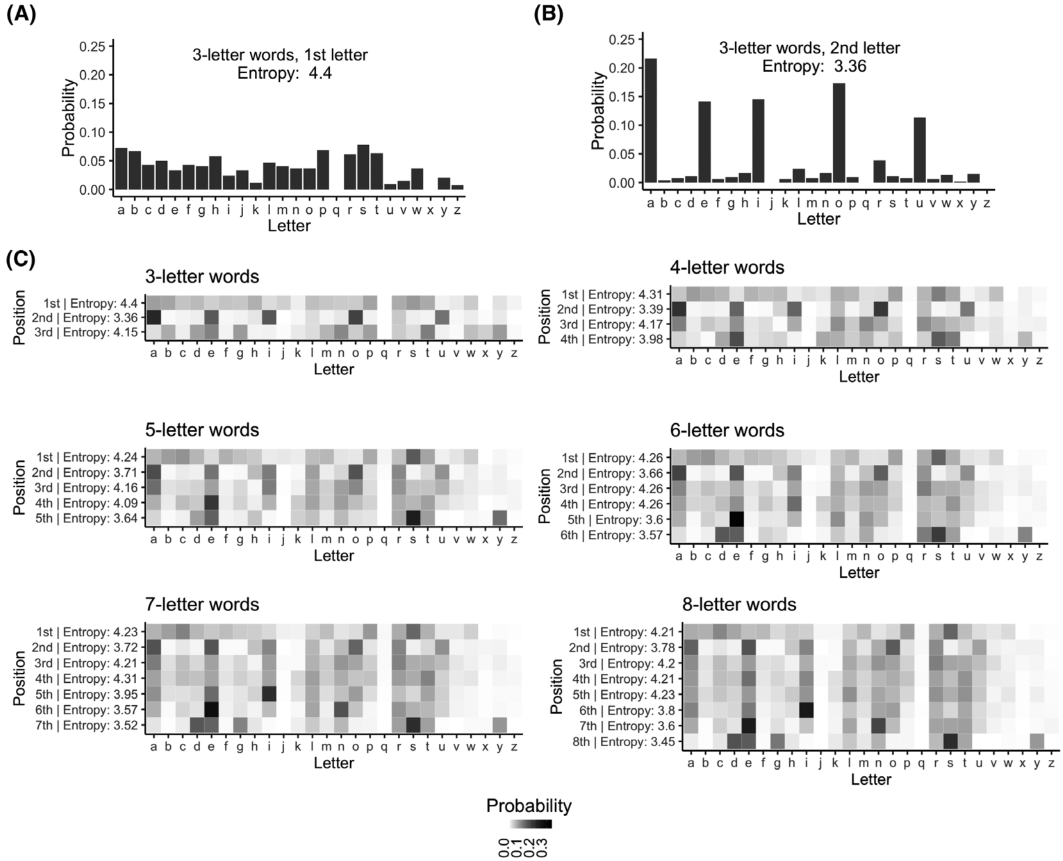

Fig. 2. We can express the a priori amount of uncertainty regarding letter identity in terms of *entropy*. In panels (A) and (B), bar plots represent the probability of each of 26 letters in specific positions for three-letter words. In panel (C), these probabilities are represented more compactly. For example, in the top left diagram in (C), the top row corresponds to the bar chart in panel (A) and the second row to panel (B). In the diagrams in (C), darker cells indicate higher probability. Row labels include entropy; a low entropy value corresponds to probability being amassed on a small number of letters, whereas high entropy indicates probability distributed across many letters.

words with at least one orthographic neighbor were considered in these analyses, as words that have zero neighbors (e.g., *ebb*, *emu*) would have zero friends and zero neighbors at every position. Thus, a total of 13,045 words were considered in these regression analyses. Log-normalized word frequencies from the HAL corpus (Lund & Burgess, 1996) were included in a nuisance regressor in each model; frequency has been shown to be a strong predictor of visual word recognition latencies (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Yap & Balota, 2009) and to modulate neighborhood effects (Andrews, 1997; Sears, Hino, & Lupker, 1995). Each model also included random intercepts for each subject and for each word. Models were implemented using mixed effects linear regression in R (R Core Team, 2018) using the *lmer* function in the "lme4" package (Bates, Maechler, Bolker, & Walker, 2015).

Results are summarized in Tables 1–6; regression estimates coming directly from the model are noted as unweighted b estimates. The critical regression estimates are denoted with E_p and indicate the influence of having relatively more enemies at a particular position p. Note that positive b estimates indicate increasingly inhibitory enemy effects on response time. There is considerable variability in the b estimates, indicating that the number of enemies has a more pronounced influence on reaction times at some letter positions than at others; that is, the deleterious influence of enemies is not constant across letter positions.

Given that there are indeed positional differences in the strength of orthographic neighborhood effects, we now turn to a consideration of whether the strength of such effects can be predicted by the a priori amount of uncertainty (i.e., entropy) associated with each letter position. Specifically, we sought to examine whether there was a significant correlation between the b estimates for each position and positional entropy values. In conducting this correlation analysis, we weighted each b value by its associated standard error as well as by how many observations contributed to the regression analysis (see Appendix A.2); weighted b values are provided in Tables 1–6. In this way, less certain b estimates would not be weighted as heavily in the correlation analysis.

The relationship between positional entropy and weighted b estimates is plotted in Fig. 3. Each point is labeled first with the word length, then with position number; for instance, the label 4.2 indicates values related to the second letter position in a four-letter word. We observed a significant negative correlation between these metrics, $r = -.482$, $t(31) = -3.064$, $p = .004$. This effect is still significant even once first-position values are excluded, $r = -.487$, $t(25) = -2.788$, $p = .010$.

2.3. Discussion

Orthographic neighborhood effects are robustly observed in the literature on visual word recognition (e.g., Andrews, 1989, 1992, 1997; Balota et al., 2007; Carreiras, Perea, & Grainger, 1997; Forster & Shen, 1996; Sears et al., 1995; Yap & Balota, 2009), such that words from larger orthographic neighborhoods are generally recognized more quickly than words from smaller neighborhoods. Critically, this measure does not distinguish

Table 1
Regression analysis for speeded word naming of three-letter words

	b Estimate (unweighted)	b Estimate (weighted)	*SE*	t Value	p Value
Intercept	6.554	—	0.018	368.280	<.001
Log Freq.	−0.013	—	0.002	−8.189	<.001
E_1	−0.004	−36.48	0.001	−3.818	<.001
E_2	−0.002	−4.07	0.002	−0.857	.392
E_3	−0.001	−2.15	0.002	−0.344	.731

Note In all tables, E_p indicates the number of enemies at letter position p.

Table 2
Regression analysis for speeded word naming of four-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	6.552	—	0.010	634.329	<.001
Log Freq.	−0.013	—	0.001	−13.871	<.001
E_1	−0.006	−184.58	0.001	−9.000	<.001
E_2	0.002	9.51	0.002	1.223	.221
E_3	−0.004	−47.70	0.001	−3.744	<.001
E_4	0.002	24.13	0.001	2.041	.041

Table 3
Regression analysis for speeded word naming of five-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	6.623	—	0.009	709.156	<.001
Log Freq.	−0.019	—	0.001	−23.811	<.001
E_1	−0.011	−408.15	0.001	−10.903	<.001
E_2	0.005	47.67	0.002	2.652	.008
E_3	−0.003	−35.72	0.002	−1.605	.109
E_4	−0.002	−23.43	0.002	−1.116	.265
E_5	0.006	42.92	0.002	2.594	.010

Table 4
Regression analysis for speeded word naming of six-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	6.636	—	0.009	721.429	<.001
Log Freq.	−0.019	—	0.001	−24.081	<.001
E_1	−0.013	−562.46	0.001	−11.189	<.001
E_2	0.003	27.84	0.002	1.175	.240
E_3	−0.006	−69.62	0.002	−2.751	.006
E_4	−0.004	−44.96	0.002	−1.702	.089
E_5	0.008	28.75	0.004	2.014	.044
E_6	−0.008	−65.48	0.003	−3.087	.002

between neighbors that mismatch the target word in one position and neighbors that mismatch the target word in another. Motivated by research suggesting that letter positions may be processed differently depending on their information content (Blais et al., 2009; Grainger & Jacobs, 1993; O'Regan et al., 1984; Pugh et al., 1994; Stevens & Grainger, 2003), we considered whether orthographic neighborhood effects differ across letter positions. Specifically, we characterized neighbors as *enemies* of the target word at every position where they mismatched and *friends* at the position where they matched. We

Table 5
Regression analysis for speeded word naming of seven-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	6.642	—	0.010	684.168	<.001
Log Freq.	−0.020	—	0.001	−22.573	<.001
E_1	−0.011	−427.36	0.001	−7.684	<.001
E_2	0.003	22.05	0.003	0.829	.407
E_3	−0.006	−31.78	0.004	−1.472	.141
E_4	0.000	−3.68	0.003	−0.131	.896
E_5	0.011	36.32	0.005	2.198	.028
E_6	0.015	22.94	0.007	2.052	.040
E_7	0.000	−2.98	0.003	−0.122	.903

Table 6
Regression analysis for speeded word naming of eight-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	6.666	—	0.011	603.976	<.001
Log Freq.	−0.021	—	0.001	−19.347	<.001
E_1	−0.021	−290.86	0.004	−5.367	<.001
E_2	0.015	88.18	0.006	2.487	.013
E_3	0.003	17.07	0.006	0.492	.623
E_4	0.006	24.60	0.007	0.825	.410
E_5	−0.007	−33.24	0.006	−1.042	.298
E_6	0.013	35.42	0.009	1.491	.136
E_7	0.015	38.81	0.009	1.687	.092
E_8	0.012	112.84	0.005	2.566	.010

observed that the deleterious influence of the number of enemies was not constant across letter positions. Moreover, we found that the influence of enemies on response times was predicted by the amount of *entropy* (i.e., uncertainty about letter identity) at each letter position. In contrast to a previous study considering positional effects in visual word recognition (Blais et al., 2009), our entropy measure was conditioned only on word length, meaning that it does not assume serial processing of the input (Coltheart & Rastle, 1994; Whitney, 2008).

We found that the inhibitory influence of enemies is most pronounced at positions with low entropy. That is, if there is low a priori uncertainty about what letter will occur at some position, each additional enemy serves to increase uncertainty about the letter identity at that position, rather than reducing uncertainty. Instead of considering these effects in terms of enemies, one can also reframe these results in terms of friends. (As shown in Appendix A.1, it is relatively straightforward to take regression values from the analyses of enemies and reexpress them to relate to friends.) Such a consideration indicates that the benefit of each additional friend is particularly pronounced at low-entropy positions,

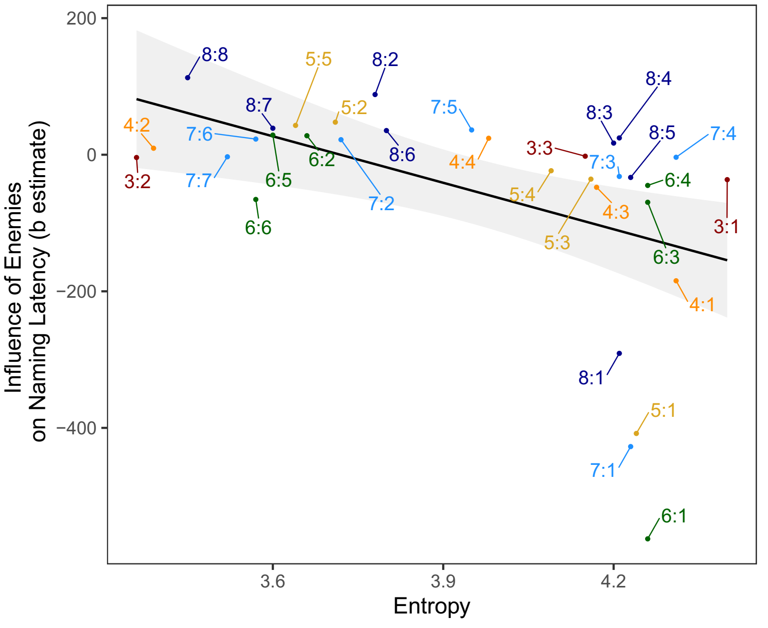

Fig. 3. The influence of enemies on naming latencies is predicted from the amount of a priori uncertainty about letter identity (entropy); specifically, enemies have the strongest inhibitory effects at low-entropy positions. Positional entropy values are plotted along the x-axis, and weighted b estimates are plotted along the y-axis. Point labels (e.g., 3:2) indicate first word length (three letters) and then position number (second position); points of the same color have the same word length.

where probability tends to be amassed on few letters. That is, friends are particularly facilitative when they confirm the identity of a letter in positions where there is little prior uncertainty about letter identity.

The present results are useful to consider in conjunction with a broader literature suggesting that readers are sensitive to differences in the information content of each letter position. Particularly interesting is a suggestion by Blais et al. (2009) that visual word recognition entails either a serial processing strategy in which positions are processed in order of their importance or a partially parallel strategy in which only letters in more important positions are processed simultaneously. It seems possible that the positional effect observed in the current analyses—namely, that there are relatively pronounced facilitation effects of friends in low-entropy places and relatively pronounced inhibitory effects of enemies in these same positions—could emerge in either such architecture. If readers prioritize extracting information from letter positions where there is high prior uncertainty about letter identity, as Blais et al. suggest, then friends may be particularly helpful at letter positions that are not being prioritized, where they can support a reader's predictions about letter identity; the strong support of these friends may facilitate bottom-up feature extraction information from high-entropy positions. Enemies are particularly injurious in such positions because they work to disconfirm a reader's predictions about letter identity. Alternatively, the positional effects observed here might also emerge in a fully parallel processing system for the simple reason that low-entropy positions will tend to have more friends, resulting in greater lexical feedback to these positions and thus position-specific benefits. In order to investigate the computational mechanisms through which the observed positional entropy effects might emerge, we next introduce a computational model of word naming, VOISeR, which considers all letter inputs in parallel. Of interest is whether these same positional entropy effects will emerge in such an architecture.

3. Word naming (computational model)

To probe the computational mechanisms that might underlie the positional entropy effects observed in human performance data, we developed a computational model of word naming known as VOISeR. VOISeR takes orthography as its input and has access to the entire input word at all time steps (i.e., its architecture does not assume serial sampling of letters). Its task is to produce an over-time sequence of phonemes, such that the model is "reading aloud" each word that is given to it.

3.1. Model construction

The structure of VOISeR is schematized in Fig. 4. In brief, the model consists of an orthographic input layer, a hidden layer, and a phonological output layer. For each word, there are 13 time steps, corresponding to the maximal number of phonemes in a word. Orthographic inputs are represented using a vector that is "clamped on" across all time steps. As there are 12 letter positions (the maximum word length) with 27 possible elements at each position (one of the 26 letters or a blank element), orthographic input is represented as a 324-unit (i.e., 12×27 elements) vector. Orthographic input is mapped via bottom-up connections to a hidden layer with 400 elements, the activation of which differs across time steps. There are also feedforward connections to an output phonological layer. At each time step, the model produces as output a 21-element vector, where each element corresponds to a linguistically principled phonetic feature (e.g., voiced, syllabic, nasal). Together, the 21 features are used to create phonemic target patterns, such that the task is to produce the series of phonemes that comprise the pronunciation of the orthographic input. To accommodate variability in the number of phonemes in our words, we also defined a target pattern (a vector of all 1s) corresponding to a "null" phoneme. At the first time step, the target pattern is the first phoneme; at the second step, the target is the second phoneme, etc. Following the final phoneme, the target pattern for subsequent steps corresponded to the null phoneme. Thus, for the input *CAT*, the output is a 13-step sequence of the features corresponding to /k/, then /æ/, then /t/, and then 10 null phonemes. In addition to its feedforward connections, VOISeR also includes top-down feedback from the phonological layer to the hidden layer, and every hidden unit is connected to every other hidden unit through recurrent connections.

To read out the “pronunciation” from VOISeR at each time step, we calculated the cosine similarity between the output phonetic feature vector and phonetic feature vectors that were defined for each phoneme; note that a unique phonetic feature vector was defined for each phoneme. We identified the phoneme vector that was most similar to the pattern produced by VOISeR, and this phoneme was considered to be the one produced at that time step.

VOISeR was trained on the full set of 37,610 words named by subjects in the ELP word naming task. The model was trained for 10,000 epochs using backpropagation through time (Werbos, 1988); this number of epochs was selected because after 10,000 epochs, the model reached a human-like level of accuracy and showed reasonable correspondence with human reaction time data (see Section 3.2). Following an approach similar to that of Seidenberg and McClelland (1989), the probability p of a word being presented on a particular epoch was proportional to its written frequency F in English (Lund & Burgess, 1996), where frequency is the number of occurrences in a corpus of approximately 131 million words. In this way, the model received more experience with high-frequency words. For our model, we used the specific formula $p = 0.05 \times \ln(F) + 0.1$. After training, the model was tested on its ability to produce each word in the lexicon.

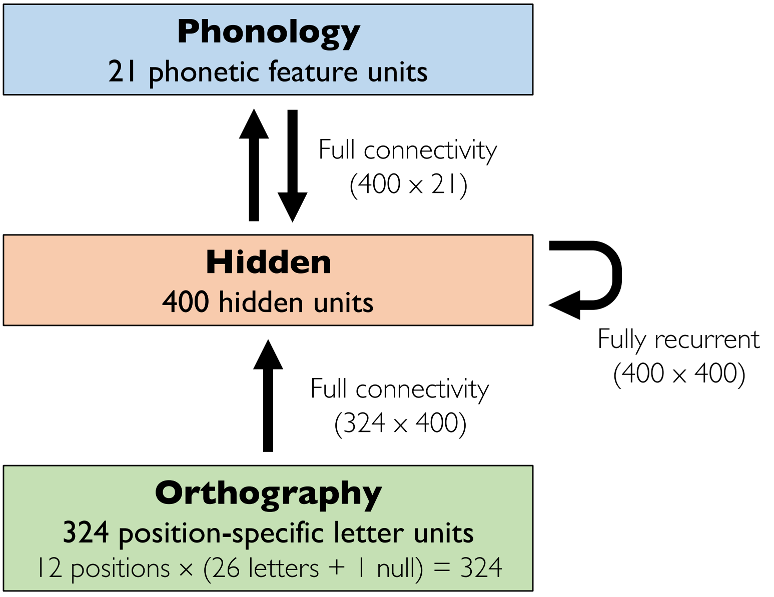

Fig. 4. A novel model of word naming, VOISeR, was used to probe the computational basis of the positional entropy effects in human subject data. VOISeR takes orthography as input, and the whole word is presented at each time step. Input is mapped to a hidden layer and ultimately to a phonological output layer, with the model producing a single phoneme at each time step. Note that the model includes top-down feedback from the phonological layer to the hidden layer, consistent with interactive frameworks, as well as recurrent connections in the hidden layer.

3.2. Model validation

Before considering whether our model can account for the differential influence of enemies at each letter position, we present the results of a few benchmark tests to validate the use of VOISeR as a model of human word naming. Specifically, we (a) compared the overall accuracy and response time profile of the model to human performance data, (b) tested for effects of frequency and orthographic neighborhood effects on overall response times from the model, and (c) examined whether the model showed human-like sensitivity to morphological consistency and regularity in its naming of words and nonwords.

In contrast to most models of visual word recognition, VOISeR is notable for the fact that it is a learning model as well as the fact that it produces an over-time sequence of phonemes as its output (i.e., it attempts to capture the serial nature of production). While additional tests would be needed to understand how VOISeR compares to leading models of visual word recognition, our goal is simply to use the model as a tool to understand how the positional entropy effects observed in the human subjects' data might emerge. We therefore defer a formal evaluation of VOISeR, noting it as an important avenue for future research.

3.2.1. Overall accuracy and response latencies

Accurate responses were ones in which the read-out pronunciation matched the target pronunciation (concatenating across all time steps); the model had an overall accuracy of 88.2%, comparable to the human subjects' overall accuracy of 88.9%. Response time was operationalized as the natural log-transformed cross entropy between the output vector produced by the model and the target output vector (again, concatenating across all time steps); there was a significant positive correlation between the response times of human subjects and those of the model, $r = .335$, $t(31{,}621) = 63.199$, $p < .001$, indicating that this output measure can serve as a proxy for response time.

3.2.2. Effects of frequency and orthographic neighborhood size

It is well established that word naming latencies in humans are sensitive to factors such as word frequency and orthographic neighborhood size (Balota et al., 2004). As such, we verified that our measure of response time showed the expected facilitative effects of word frequency, $r = -.289$, $t(31{,}621) = -53.646$, $p < .001$, and of neighborhood size, $r = -.383$, $t(31{,}621) = -73.662$, $p < .001$.

3.2.3. Consistency and regularity effects

Many researchers (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 2004; Perry, Ziegler, & Zorzi, 2007) have suggested that an important benchmark for computational models of visual word recognition is that they can account for the attested sensitivity of human readers to the consistency and regularity of grapheme–phoneme correspondences (Glushko, 1979; Taraban & McClelland, 1987). *Consistent* words are those containing letter strings like *–ink*, which is always pronounced the same way. Inconsistent words are ones that contain sequences like *–int*, which is usually pronounced as [ɪnt] as in *mint* but occasionally is pronounced as [ɑɪnt] as in *pint*. For brevity, we refer to inconsistent, regularly pronounced words (e.g., *mint*) as simply *inconsistent*, and inconsistent, irregularly pronounced words (e.g., *pint*) as *exception* words. Previous studies have established that readers are sensitive to these differences in both word and nonword naming, with the fastest, most accurate responses for consistent items, intermediate performance on inconsistent items, and worst performance on exception items. Of interest is whether these same patterns emerge in simulated data from VOISeR.

To this end, we examined the trained model's performance on the consistent, inconsistent, and exception words used by Taraban and McClelland (1987). VOISeR showed equivalent accuracy on consistent (*M*: 97.9%, *SE*: 2.1%) and inconsistent (*M*: 97.9%, *SE*: 2.1%) words, but worse performance on exception words (*M*: 75.0%, *SE*: 6.3%) words. To test whether accuracy depended on word type, we implemented a mixed effects model using the *mixed* function in the "afex" package (Singmann, Bolker, Westfall, & Aust, 2018); this function implements the *lmer* function used for other mixed effects analyses but reports the results in an ANOVA-like output, making it particularly useful when there are more than two levels of a factor. Our model included a fixed factor of Type and random intercepts for each word. This analysis indicated that there was a statistically significant difference in accuracy as a function of word type, $\chi^2(2) = 18.43$, $p < .001$. We also examined our proxy for response latencies, limiting our analysis to accurately pronounced items. VOISeR showed the fastest responses to consistent words (*M*: −8.070, *SE*: 0.193), intermediate responses to inconsistent words (*M*: −7.358, *SE*: 0.158), and slowest responses to exception words (*M*: −6.158, *SE*: 0.148); note here that smaller (i.e., more negative) cross-entropy values correspond to faster response times. Results of a linear regression indicated that all three pairwise differences were statistically significant (consistent vs. inconsistent: $b = 0.712$, $SE = 0.233$, $t = 3.052$, $p = .003$; inconsistent vs. exception: $b = 1.200$, $SE = 0.251$, $t = 4.787$, $p < .001$; consistent vs. exception: $b = 1.912$, $SE = 0.251$, $t = 7.630$, $p < .001$). Thus, VOISeR's recognition of words shows sensitivity to the consistency and regularity of grapheme–phoneme correspondences, as is also observed in human subjects.

Finally, we tested for consistency effects in naming of nonwords like *bink* and *bint*. For these analyses, we employed nonword stimuli previously used by Glushko (1979); acceptable pronunciations for each item were taken from Plaut, McClelland, Seidenberg, and Patterson (1996). Glushko found that human subjects were more accurate in naming consistent nonwords (*M*: 93.8%) than inconsistent nonwords (*M*: 78.3%). Similarly, VOISeR showed numerically greater accuracy on consistent nonwords (*M*: 81.4%, *SE*: 6.0%) than inconsistent ones (*M*: 74.4%, *SE*: 6.7%); however, this difference was not significant, $\chi^2(1) = 0.61$, $p = .43$, as per an analysis that followed the same approach as in the analysis of word data. Finally, we tested for consistency effects in response time data, following the same approach as in the analysis of the word data. VOISeR showed significantly faster response times for consistent nonwords (*M*: −7.135, *SE*: 0.240) than inconsistent nonwords (*M*: −6.484, *SE*: 0.184), $b = 0.650$, $SE = 0.307$, $t = 2.123$, $p = .038$. Thus, VOISeR's performance shows the expected influence of phoneme–grapheme consistency on nonword recognition.

Taken together, the analyses presented in this section suggest that VOISeR is a valid computational model of visual word recognition and motivate its use in understanding the mechanisms through which positional entropy effects might emerge. We turn next to the results of simulations considering whether VOISeR approximates the friend/enemy effects observed in the human subject performance data.

3.3. Results

To assess the influence of the number of enemies at each letter position, we conducted a series of regression analyses similar to those conducted for the human performance data. As before, a separate analysis was conducted for each word length. The dependent variable was our measure of response time (natural log-transformed cross entropy), and the model included a regressor for each letter position. Only words with at least one orthographic neighbor contributed to these analyses. Each model also included a nuisance regressor for log-transformed word frequency (Lund & Burgess, 1996) values. In contrast to the analysis of human subjects' data, the data here do not come from multiple subjects, and there is only one observation per word in this data set; as such, we do not estimate random intercepts for subject or item. Thus, models were implemented using the *lm* function in the "stats" package (R Core Team, 2018).

Results are summarized in Tables 7–12. Note that as before, we report unweighted *b* values (i.e., those that come directly from the regression analyses) as well as weighted *b* values, which have been weighted in accordance with the associated standard error (see Appendix A.2 for details).

As in the human subjects' data, there is a considerable amount of variability in the estimated regression estimates, indicating that the influence of enemies is not constant across letter position. We thus examined whether weighted regression estimates were predicted by the a priori amount of uncertainty (i.e., entropy) about letter identity at each position. As illustrated in Fig. 5, there was a significant negative correlation between these metrics, $r = -.413$, $t(31) = -2.32$, $p = .017$, indicating that the inhibitory influence of enemies is most pronounced in low-entropy positions. Visually, this correlation appears to be driven primarily by word-initial positions; note that the weighted *b* estimates for these positions are relatively large because the regression estimates are associated with relatively little standard error. In contrast to the analyses of data from the human subjects,

Table 7
Regression analysis for VOISeR's word naming of three-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	−6.271	—	0.316	−19.835	<.001
Log Freq.	−0.008	—	0.030	−0.264	.792
E_1	−0.203	−55.736	0.021	−9.916	.000
E_2	0.029	1.758	0.043	0.659	.510
E_3	−0.104	−12.475	0.031	−3.350	.001

Table 8
Regression analysis for VOISeR's word naming of four-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	−6.447	—	0.145	−44.358	<.001
Log Freq.	−0.045	—	0.016	−2.719	.007
E_1	−0.094	−85.389	0.012	−7.743	<.001
E_2	−0.079	−10.504	0.032	−2.481	.013
E_3	−0.066	−23.321	0.019	−3.384	.001
E_4	−0.087	−26.691	0.021	−4.173	<.001

Table 9
Regression analysis for VOISeR's word naming of five-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	−6.449	—	0.110	−58.454	<.001
Log Freq.	−0.059	—	0.014	−4.311	<.001
E_1	−0.110	−131.301	0.016	−6.704	<.001
E_2	−0.163	−46.514	0.034	−4.852	<.001
E_3	−0.037	−15.328	0.028	−1.322	.186
E_4	−0.117	−45.175	0.029	−4.044	<.001
E_5	−0.077	−18.092	0.037	−2.083	.037

Table 10
Regression analysis for VOISeR's word naming of six-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	−6.308	—	0.093	−68.144	<.001
Log Freq.	−0.054	—	0.012	−4.416	<.001
E_1	−0.177	−253.852	0.017	−10.201	<.001
E_2	−0.286	−92.067	0.037	−7.799	<.001
E_3	−0.078	−27.883	0.035	−2.236	.025
E_4	−0.111	−41.606	0.034	−3.263	.001
E_5	−0.180	−19.780	0.063	−2.874	.004
E_6	−0.053	−13.003	0.042	−1.257	.209

the relationship between the b estimates and the entropy values is no longer significant once word-initial positions are excluded from analyses, $r = -.216$, $t(25) = -1.11$, $p = .278$.

Finally, we considered how well the weighted b estimates from the analyses of VOISeR's performance compared to the weighted b estimates from the analyses of human naming data. As shown in Fig. 6, there is a strong correlation between these two sets of

Table 11
Regression analysis for VOISeR's word naming of seven-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	−6.455	—	0.092	−70.337	<.001
Log Freq.	−0.052	—	0.013	−4.129	<.001
E_1	−0.206	−259.405	0.021	−9.996	<.001
E_2	−0.259	−77.620	0.042	−6.135	<.001
E_3	−0.233	−46.258	0.052	−4.488	<.001
E_4	−0.040	−13.260	0.040	−0.992	.321
E_5	−0.142	−16.300	0.068	−2.084	.037
E_6	0.254	12.207	0.105	2.411	.016
E_7	−0.174	−43.343	0.046	−3.762	<.001

Table 12
Regression analysis for VOISeR's word naming of eight-letter words

	b Estimate (unweighted)	*b* Estimate (weighted)	*SE*	*t* Value	*p* Value
Intercept	−5.963	—	0.105	−56.577	<.001
Log Freq.	−0.090	—	0.014	−6.509	<.001
E_1	−0.405	−185.611	0.048	−8.414	<.001
E_2	−0.425	−88.346	0.072	−5.947	<.001
E_3	−0.450	−89.397	0.073	−6.153	<.001
E_4	−0.323	−45.986	0.086	−3.740	<.001
E_5	−0.348	−57.462	0.080	−4.337	<.001
E_6	0.007	0.617	0.107	0.062	.950
E_7	0.162	12.610	0.117	1.387	.165
E_8	−0.134	−41.973	0.058	−2.301	.022

regression estimates, $r = .834$, $t(31) = 8.42$, $p < .001$; however, this is driven almost entirely by word-initial positions, and the correlation is nonsignificant once these positions are excluded, $r = -.169$, $t(25) = -.85$, $p = .401$. As before, this is likely because the unweighted *b* estimates for word-initial positions are associated with relatively little standard error; as such, weighting these estimates by standard error leads to relatively large values for the word-initial positions and shrinks the range of the estimates for the other letter positions, potentially making it difficult to see correlations with the *b* estimates from the human naming data.

3.4. Discussion

Analyses of human subject performance in a speeded word naming task indicated that the positional influence of enemies on response times was predicted by the a priori degree of uncertainty about letter identity at that letter position (Section 2). In Section 3, we introduced a novel model of word naming, VOISeR, to probe the computational

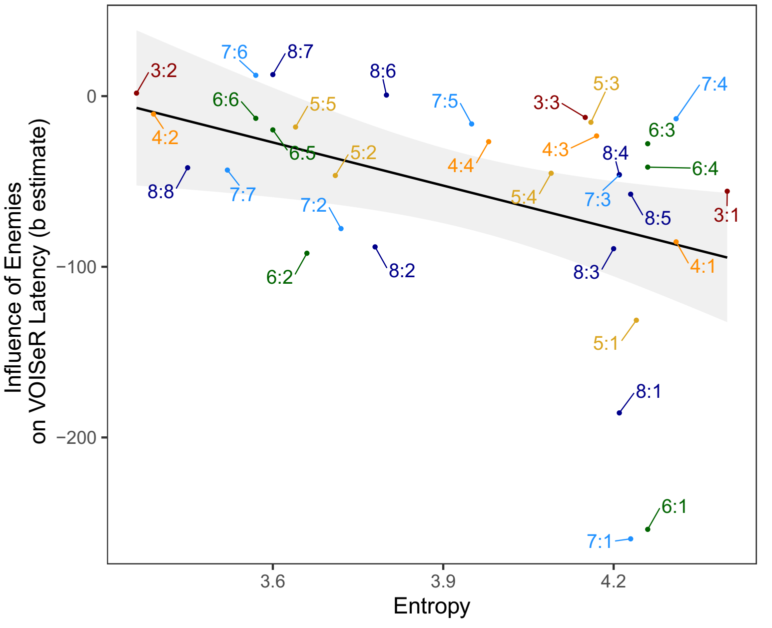

Fig. 5. The influence of enemies on simulated word naming times is predicted from the amount of a priori uncertainty about letter identity (entropy); enemies have the strongest inhibitory effects on VOISeR's performance at low-entropy positions. However, this correlation is not significant once word-initial positions are excluded.

mechanisms through which such positional entropy effects might have emerged. VOISeR, which receives orthographic inputs in parallel and produces an over-time sequence of phonemes, showed reasonable correspondence with human subjects' performance, naming words with comparable accuracy (model: 88.2%, humans: 88.9%) and overall response times (r = .335). As in the human subjects' data, regression analyses indicated that the influence of enemies was not constant across letter position and was predicted by positional entropy, with the inhibitory effect of each additional enemy being most pronounced in low-entropy positions. Notably, the relationship between entropy and the enemy regression estimates was driven primarily by word-initial positions. Furthermore, any similarity between the b values obtained from the VOISeR analyses and the b values obtained from the human subjects' data was driven almost entirely by word-initial positions. Taken together, these results indicate that VOISeR's ability to capture positional entropy effects may be limited to the first letter position of a word. Thus, while VOISeR can partially approximate positional entropy effects observed in human performance data, its ability to do so may be limited to word-initial positions. In general, the first letter position tends to be particularly informative for visual word recognition (e.g., Grainger, O'Regan, Jacobs, & Segui, 1992; Inhoff & Tousman, 1990; O'Regan et al., 1984).

These computational simulations suggest that positional entropy effects can emerge, at least to some degree, in a cognitive system where orthographic inputs are sampled in a

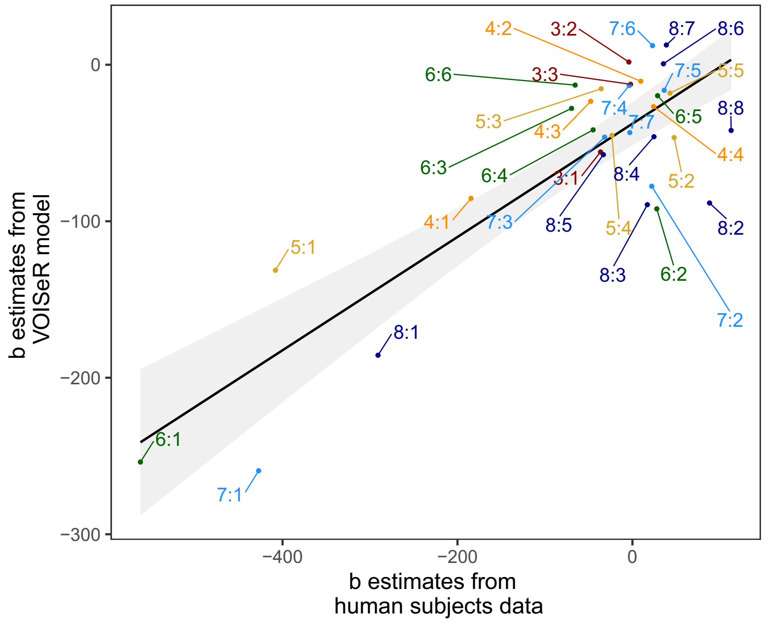

Fig. 6. Weighted *b* estimates from VOISeR are correlated with weighted *b* estimates from human subjects' data, but the correlation is almost entirely driven by word-initial positions.

purely parallel fashion. This is in contrast to the suggestion made by Blais et al. (2009), who suggested that visual word recognition relies either on a serial processing mechanism or on one that is only partly parallel. Thus, the present simulations suggest that positional entropy effects may emerge due to internal processing dynamics, rather than due to positional differences in input sampling. Specifically, the recurrent connections within the model may allow VOISeR to more heavily weight some positions than others, effectively "attending" more to some positions than others.

So far, we have provided both behavioral and computational evidence to describe how neighborhood effects differ across letter positions during visual word identification and have demonstrated that entropy offers a useful way to characterize the strength of these effects. Still, it is worth remarking that our entropy metric considers only orthography. Speeded naming tasks require mapping from orthography to phonology, and it is striking that the letters that are most probable in low-entropy positions tend to be ones with highly irregular grapheme–phoneme mappings (e.g., vowels). In other words, positional differences in entropy may be confounded with differences in phoneme–grapheme consistency, at least in English. As such, it may be the case that the positional effects observed in the current work reflect not the information content of the letter position but rather the likelihood that the position contains a letter that is easy to map to phonology. The notion that the relative importance of different letter positions may arise from the orthography-to-phonology mapping process is in line with a literature examining the relationship between orthography and phonology in visual word recognition. For instance, work by

Adelman and Brown (2007) suggests that neighborhood effects may be a consequence of print-sound conversion processes rather than as a result of top-down, bottom-up, and lateral interactions among the word and letter layers of the orthographic system (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981). In order to probe the role of print-sound conversion in the position-specific neighborhood effects found in word naming, we ask next whether such effects might emerge in a lexical decision task, which does not explicitly require mapping to phonology.

4. Lexical decision

While both word naming and lexical decision tasks are often used to study visual word recognition, there are important differences between such tasks. Word naming requires readers to explicitly map to phonology and also requires additional steps beyond recognition per se (e.g., articulatory planning). Lexical decision, by contrast, may not require a reader to explicitly select a single lexical item, as readers can conceivably make their decisions based on how word-like a stimulus is (Andrews, 1997; Grainger & Jacobs, 1996). More generally, the influence of lexical metrics can differ across visual word recognition tasks (Balota & Chumbley, 1984), though larger neighborhoods do tend to be associated with faster visual word recognition in both speeded naming and lexical decision tasks (Andrews, 1997). As such, we also examined whether the entropy-based friend/enemy effects that emerged in word naming would also emerge in a lexical decision task.

4.1. Methods

Trial-by-trial lexical decision data were obtained from the ELP (Balota et al., 2007). As in the word naming task, analyses were limited to words that were three to eight letters in length. This resulted in 23,701 total words, sampled across 797 participants. We only analyzed data from trials in which participants accurately classified word stimuli as words, yielding an average of 28.5 ($SD = 6.9$) observations per item. Analyses considered the number of enemies per letter position and the entropy value associated with each position, which were calculated as described in Section 2.1.

4.2. Results

Separate regression analyses were conducted for each level of word length. Natural log-transformed reaction times were used as the dependent variable, and each analysis included separate regressors for the number of enemies at each letter position as well as a nuisance regressor for log-transformed word frequencies (Lund & Burgess, 1996). By-subject and by-item random intercepts were also modeled. As in Section 2, analyses were implemented using the *lmer* function in the "lme4" package (Bates et al., 2015).

Results are presented in Tables 13–18. As before, *b* estimates are not constant across letter position. We thus considered how *b* values related to positional entropy values, using weighted b values in our correlation so that each b estimate was weighted according to its associated level of certainty (Appendix A.2). As shown in Fig. 7, there was a significant negative correlation between entropy values and the weighted b values, $r = -.423$, $t(31) = -2.601$, $p = .014$. This may be driven primarily by first-position values, as the correlation was no longer significant once these positions were excluded, $r = -.209$, $t(25) = -1.066$, $p = .297$.

4.3. Discussion

Analyses of lexical decision data indicate that as in word naming, the influence of orthographic neighbors is not constant across letter positions; rather, the inhibitory influence of enemies (and correspondingly, the facilitative influence of friends) is most pronounced in low-entropy positions, which are associated with relatively little a priori uncertainty about letter identity. The finding that entropy-mediated friend/enemy effects emerge in both tasks could suggest that such effects relate to a core part of visual word recognition, rather than a peripheral task-specific process (such as articulatory planning in word naming or decision-level processes in lexical decision). However, it is also possible that the presence of positional entropy effects in both tasks is due to different mechanisms; for instance, the effects in word naming may be driven by positional differences in the consistency of grapheme–phoneme correspondences, whereas the effects in lexical decision may be driven by positional differences in uncertainty about letter identity, independent of pronunciation. To test whether the positional effects reported here are driven by entropy per se, it might be informative to assess whether these effects generalize across languages with varying orthographic depth (i.e., languages that vary in the consistency of grapheme–phoneme correspondences). If the positional effects seen here are also truly driven by entropy, they should emerge in other languages, too, regardless of orthographic depth.

5. General discussion

In general, recognition of a visual target word is facilitated by the presence of orthographically similar words in the lexicon, as shown both in speeded word naming and

Table 13
Regression analysis for lexical decision of three-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	6.785	—	0.024	284.753	<.001
Log Freq.	−0.034	—	0.002	−15.538	<.001
E_1	−0.007	−65.059	0.001	−4.901	<.001
E_2	−0.005	−11.973	0.003	−1.797	.073
E_3	0.005	18.117	0.002	2.086	.038

Table 14
Regression analysis for lexical decision of four-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	6.767	—	0.011	629.483	<.001
Log Freq.	−0.035	—	0.001	−36.161	<.001
E_1	−0.004	−115.778	0.001	−5.321	<.001
E_2	−0.003	−14.712	0.002	−1.784	.075
E_3	−0.004	−46.201	0.001	−3.417	.001
E_4	−0.003	−29.587	0.001	−2.367	.018

Table 15
Regression analysis for lexical decision of five-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	6.785	—	0.009	733.986	<.001
Log Freq.	−0.038	—	0.001	−46.696	<.001
E_1	−0.005	−219.060	0.001	−5.688	<.001
E_2	−0.007	−60.669	0.002	−3.285	.001
E_3	−0.003	−48.582	0.002	−2.126	.034
E_4	−0.007	−94.771	0.002	−4.385	<.001
E_5	−0.009	−67.666	0.002	−3.969	<.001

Table 16
Regression analysis for lexical decision of six-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	6.782	—	0.009	760.804	<.001
Log Freq.	−0.037	—	0.001	−47.390	<.001
E_1	−0.003	−142.911	0.001	−2.767	.006
E_2	−0.007	−72.697	0.002	−2.977	.003
E_3	−0.002	−24.817	0.002	−0.957	.339
E_4	−0.005	−64.399	0.002	−2.362	.018
E_5	−0.003	−12.404	0.004	−0.843	.399
E_6	−0.009	−73.206	0.003	−3.348	.001

lexical decision tasks (e.g., Balota et al., 2004). These orthographic neighbors are typically defined as words that mismatch the target word at a single letter position (Coltheart et al., 1977), but crucially, this definition does not consider *where* neighbors mismatch the target word. This is noteworthy because not all letter positions are equally informative with regard to word identity. In some positions, probability tends to be well distributed among many possible letter identities; such positions are referred to as having *high*

Table 17
Regression analysis for lexical decision of seven-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	6.769	—	0.009	726.104	<.001
Log Freq.	−0.035	—	0.001	−40.950	<.001
E_1	−0.004	−159.610	0.001	−2.818	.005
E_2	−0.006	−54.240	0.003	−1.996	.046
E_3	−0.001	−6.897	0.004	−0.314	.753
E_4	−0.007	−73.504	0.003	−2.549	.011
E_5	−0.010	−36.818	0.005	−2.177	.030
E_6	0.016	25.718	0.007	2.267	.023
E_7	−0.011	−88.147	0.003	−3.538	<.001

Table 18
Regression analysis for lexical decision of eight-letter words

	b Estimate (unweighted)	b Estimate (weighted)	SE	t Value	p Value
Intercept	6.770	—	0.011	641.760	<.001
Log Freq.	−0.032	—	0.001	−32.028	<.001
E_1	−0.008	−130.506	0.004	−2.402	.016
E_2	−0.016	−104.437	0.005	−2.942	.003
E_3	0.002	14.013	0.005	0.403	.687
E_4	−0.012	−55.177	0.006	−1.843	.065
E_5	−0.019	−104.112	0.006	−3.241	.001
E_6	0.014	41.658	0.008	1.742	.082
E_7	0.020	55.014	0.008	2.379	.017
E_8	0.002	22.821	0.004	0.516	.606

entropy because knowing the letter identity at this position is highly informative. Low-entropy positions, by contrast, are ones where probability tends to be amassed on a small set of letters. Evidence suggests that readers are differentially sensitive to letter positions within a word (Forster & Davis, 1991; Grainger & Jacobs, 1993; Grainger et al., 1992; Inhoff & Tousman, 1990; O'Regan et al., 1984; Stevens & Grainger, 2003; Whitney, 2001) and that they may specifically be sensitive to the degree of uncertainty (i.e., the entropy) associated with each letter position (Blais et al., 2009). The notion that processing might favor letter positions with high information content would be consistent with Bayesian accounts of perception, which posit that individuals make optimal (or at least, near-optimal) inferences about the underlying structure of the perceptual input they receive (e.g., Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kleinschmidt & Jaeger, 2015; Magnuson, Mirman, Luthra, Strauss, & Harris, 2018; McClelland, 2013; McClelland, Mirman, Bolger, & Khaitan, 2014; Norris & McQueen, 2008; Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

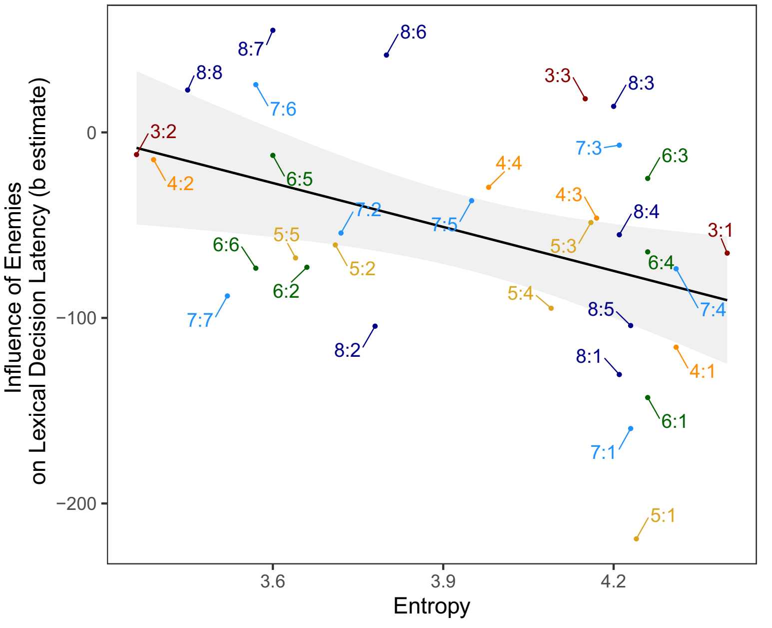

Fig. 7. The influence of enemies on latency in a lexical decision task is predicted by the degree of entropy associated with each letter position. As in word naming, the inhibitory influence of enemies is most pronounced in low-entropy positions.

The present investigation was undertaken to examine whether orthographic neighborhood effects differ across letter positions. To do so, we leveraged the notion of *friends* and *enemies* introduced by McClelland and Rumelhart (1981); a neighbor can be considered a friend to a target word at every letter position where it matches and an enemy at every position where it mismatches. Using a large set of human performance data (Balota et al., 2007), we examined how the number of enemies at each letter position predicts response latencies. Across both speeded word naming (Section 2) and lexical decision (Section 4) paradigms, we observed that the influence of enemies was not constant across letter position. Rather, enemies were particularly inhibitory in low-entropy positions, and equivalently, friends were particularly facilitative in these positions (Appendix A.1). Such a result can be understood by recognizing that low-entropy positions are associated with low uncertainty about letter identity. At such positions, each additional friend supports a reader's predictions about what letter they will encounter, potentially supporting bottom-up feature extraction in high-entropy positions where readers are less certain about letter identity. By contrast, any additional enemy at a low-entropy position serves to increase uncertainty in a position where readers are ordinarily confident about which letters they will encounter; as such, enemies are particularly destabilizing in low-entropy positions. The fact that these results were consistent across naming and lexical decision tasks may suggest that positional neighbor influences operate on a core part of the visual word recognition process, rather than some process (e.g., motor planning) that is specific to one of the two tasks.

We also present simulations from a novel computational model, VOISeR (Section 3); this model receives orthographic input and is tasked with producing an over-time sequence of phonemes. The model data show a reasonable correspondence with human accuracy and response time data, and VOISeR is also able to approximate positional entropy effects at least in word-initial positions. As such, VOISeR may provide some insight into the computational mechanisms underlying the positional entropy effects seen in human performance data. In particular, the fact that VOISeR receives its inputs in parallel suggests that position-specific effects can emerge in a model architecture that does not require serial processing of letter inputs. Instead, position-specific effects seem to arise from internal dynamics of the model, which effectively lead the model to differentially weight letter positions as a function of entropy.

In summary, the present investigation brings together work on neighborhood effects in visual word recognition with research suggesting that various letter positions may be differentially important in word identification. Our analyses indicate that the number of enemies (or friends) at a given letter position is a useful predictor of word identification latency and that the negative consequences of enemies (or, equivalently, the benefits of friends) for response latencies are most pronounced at low-entropy letter positions, where there is little a priori uncertainty about letter identity. Our novel computational model provides some insight into how such an effect may emerge, suggesting that such positional entropy effects can emerge in a system that does not sample inputs in parallel. Note that since the current work specifically tested for a relationship between Shannon entropy of a given letter position and the influence of neighbors mismatching at that position, future work is needed to understand whether other information theoretic metrics (e.g., surprisal) might better capture how neighborhood effects differ across letter positions. Overall, the present work suggests that studies of visual word recognition should carefully consider how information about word identity is distributed between various letter positions, as this can have important consequences for how the recognition process unfolds.

Acknowledgments

All analysis scripts are available online at https://github.com/sahil-luthra/friends-and-enemies. We thank Dave Kenny for assistance with data analyses as well as an anonymous reviewer for commentary on a previous version of this manuscript. We also thank attendees of the 2017 and 2018 meetings of the Cognitive Science Society as well as attendees of the 2018 meeting of the Psychonomic Society for helpful feedback on earlier versions of this project. This research was supported in part by NSF grants IGERT DGE-1144399, BCS-PAC 1754284, and NRT DGE-1747486 (PI: J.S.M.), by NIH NICHD grant P20 HD091013 (PI: J.G.R.), and by an NSF Graduate Research Fellowship to S.L.

Open Research badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://github.com/sahil-luthra/friends-and-enemies.

Notes

1. Here, we use the terms *friend* and *enemy* as they are used by McClelland and Rumelhart (1981). Note that this is distinct from another use of these terms in the literature, in which they refer to sets of words with similar or dissimilar phoneme–-grapheme correspondence, respectively (e.g., Kay & Bishop, 1987; Taraban & McClelland, 1987).
2. We note that our entropy metric generally approximates the relative importance values calculated by Blais et al. (2009), r = .490, $t(20)$ = 2.51, p = .02. Note that this correlation reflects only words that are four, five, six, or seven letters in length, as Blais et al. did not consider three- or eight-letter words.

References

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review*, *14*(3), 455–459.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(5), 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*(4), 439–461.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357.

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.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Blais, C., Fiset, D., Arguin, M., Jolicoeur, P., Bub, D., & Gosselin, F. (2009). Reading between eye saccades. *PLoS One*, *4*(7), e6448.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*(4), 857–871.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1197–1211.

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.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, *30*, 1–25.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning Memory and Cognition*, *22*(3), 696–713.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674–691.

Grainger, J., & Jacobs, A. M. (1993). Masked partial-word priming in visual word recognition: Effects of positional letter frequency. *Journal of Experimental Psychology-Human Perception and Performance*, *19* (5), 951–964.

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., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics*, *51*(1), 49–56.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720.

Inhoff, A. W., & Tousman, S. (1990). Lexical priming from partial-word previews. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 825–836.

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology*, *26*, 240–346.

Kay, J., & Bishop, D. (1987). Anatomical differences between nose, palm, and foot, or, the body in question: Further dissection of the processes of sub-lexical spelling-sound translation. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 449–469). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 203–208.

Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. D. (2018). Interaction in spoken word recognition models: Feedback helps. *Frontiers in Psychology*, *9*(369), 1–18.

Mathey, S., & Zagar, D. (2000). The neighborhood distribution effect in visual word recognition: Words with single and twin neighbors. *Journal of Experimental Psychology: Human Perception and Performance*, *26* (1), 184–205.

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: A historical and tutorial review. *Frontiers in Psychology*, *4*, 1–25.

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *38*(6), 1139–1189.

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.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillière, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 393.

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.

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.

Pugh, K. R., Rexer, K., Peter, M., & Katz, L. (1994). Neighborhood effects in visual word recognition: Effects of letter delay and nonword context difficulty. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*(3), 639–648.

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

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 876–900.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). afex: Analysis of factorial experiments. R package version 0.21-2. Available at: https://CRAN.R-project.org/package=afex

Stevens, M., & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception & Psychophysics*, *65*(1), 133–151.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, *26*(6), 608–631.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, *1*(4), 339–356.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review*, *8*(2), 221–243.

Whitney, C. (2008). Supporting the serial in the SERIOL model. *Language and Cognitive Processes*, *23*(6), 824–865.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502–529.

Appendix A: Mathematical details

A.1 Converting between friends and enemies

While our results are described in terms of the number of enemies at each position, readers may find it informative to instead consider results in terms of the number of friends. By definition, the number of friends at a given position can be computed from the number of enemies at all other positions (Fig. 1). In a three-letter word, for instance, the number of friends at position 1, denoted F_1, can be described in terms of the number of enemies at positions 2 and 3 (E_2 and E_3). That is,

$$F_1 = E_2 + E_3$$

or, equivalently,

$$F_1 = -E_1 + E_1 + E_2 + E_3 = -E_1 + \sum_{i=1}^{3} E_i.$$

More generally, for any n-letter word, the number of friends at position p can be described in terms of enemies.

$$F_p = -E_p + \sum_{i=1}^{n} E_i.$$

The resulting set of equations can be algebraically transformed such that the number of enemies at position p is expressed in terms of friends.

$$E_p = -F_p + \frac{1}{n-1}\sum_{i=1}^{n} F_i$$

The analyses reported in this paper indicate how the number of enemies at each position influences word recognition times. That is, the regression estimates (unweighted b values) are multiplied by the number of enemies at each position to compute response times. For three-letter words,

$$\ln(RT) = b_0 + b_1E_1 + b_2E_2 + b_3E_3 + \text{error}.$$

Because of the mathematical relationship between friends and enemies, this regression equation can be rewritten in terms of the number of friends at each position. Based on the equations given above, the appropriate coefficients for the number of friends at each position would simply be linear combinations of the reported b values for the number of enemies at each position. For three-letter words,

$$\ln(RT) = b_0 + \left(\frac{-b_1 + b_2 + b_3}{2}\right)F_1 + \left(\frac{b_1 - b_2 + b_3}{2}\right)F_2 + \left(\frac{b_1 + b_2 - b_3}{2}\right)F_3 + \text{error}.$$

More generally, for n-letter words,

$$\ln(RT) = b_0 + \left(-b_1 + \frac{1}{n-1}\sum_{i=1}^{n} b_i\right)F_1 + \left(-b_2 + \frac{1}{n-1}\sum_{i=1}^{n} b_i\right)F_2 + \cdots + \left(-b_n + \frac{1}{n-1}\sum_{i=1}^{n} b_i\right)F_n + \text{error}.$$

A.2 Weighting regression estimates for correlation analyses

To ensure that less certain b estimates were not weighted as heavily in the correlation analyses, we weighted the estimates obtained from each regression analysis. For instance, in the analysis of three-letter words, the b estimate for the first letter position was weighted as follows:

$$b_{E_1,\text{weighted}} = \left(b_{E_1,\text{unweighted}}\right)\left(\frac{N}{\frac{1}{(SE_1)^2}+\frac{1}{(SE_2)^2}+\frac{1}{(SE_3)^2}}\right)\left(\frac{1}{(SE_1)^2}\right),$$

where subscripts indicate the letter position, b indicates the (weighted or unweighted) regression estimate, SE indicates the standard error associated with a regression estimate, and N indicates the number of observations that contributed to the regression analysis. More generally, in an analysis of n-letter words, the b estimate for position p was weighted as follows:

$$b_{E_p,\text{weighted}} = \left(b_{E_p,\text{unweighted}}\right)\left(\frac{N}{\sum_{i=1}^{n}\frac{1}{(SE_i)^2}}\right)\left(\frac{1}{(SE_p)^2}\right).$$

In this way, less certain estimates (i.e., those associated with higher standard error) were not weighted as heavily as more certain estimates. Furthermore, the more observations that contributed to a regression analysis, the more heavily a particular b value was weighted.