



Readers target words where they expect to minimize uncertainty

Jon W. Carr^{a,b,*}, Monica Fantini^a, Lorena Perrotti^a, Davide Crepaldi^a

^a Cognitive Neuroscience, International School for Advanced Studies (SISSA), Trieste, Italy

^b Department of Psychology, Royal Holloway, University of London, England, United Kingdom

ARTICLE INFO

Keywords:

Eye movements
Optimal viewing position
Reading
Statistical learning
Visual word recognition

ABSTRACT

Skilled readers use multiple heuristics to guide their eye movements during reading. One possible cue that readers may rely on is the way in which information about word identity is typically spread across words. In many (but not all) languages, words are, on average, more informative on the left, predicting that readers should have a preference for left-of-center fixation when targeting words. Any such effect will, however, be modulated by important perceptual constraints and may be masked by various confounding factors. In three experiments with artificially constructed lexicons, we provide causal evidence that the way in which a language distributes information affects how readers land on words. We further support our analyses with a Bayesian cognitive model of visual word recognition that predicts where readers ought to fixate in order to minimize uncertainty about word identity. Taken together, our findings suggest that global properties of the lexicon may play a role in isolated word targeting, and may therefore make a contribution to eye movement behavior in more natural reading settings.

Introduction

Skilled readers are able to interpret a written text with remarkable ease and efficiency despite the fact that writing is a complex code that takes on a multitude of visual forms and reflects the many layers of complexity that are inherent to human language—from individual letters and how they combine to morphosyntactic structure and beyond. To make this possible, readers must learn reliable heuristics, rapidly combine multiple sources of information, and contend with a range of competing trade-offs (Rayner, 1998; Snowling et al., 2022; Yeatman & White, 2021).

One general cue that readers may rely on is the way in which information about word identity is typically spread across words within a given lexicon (Alhama et al., 2019; Clark & O'Regan, 1999; Deutsch & Rayner, 1999; Farid & Grainger, 1996; Shafir et al., 2022). In many—but not all—languages, the most informative content is, on average, concentrated closer to word onset. For example, the first three letters of *guarded* are much more informative than the last three, since there are few words in English that begin with *gua-* and many that end with *-ded*. A number of explanations have been put forward to explain why words might be more informative at their beginnings. The most obvious reason is because languages are—first and foremost—adapted for the efficient production and comprehension of speech; thus, there may be pressures

acting on the language to position informative content earlier in words, allowing the so-called uniqueness point to be reached more rapidly (Marslen-Wilson, 1987). Another possibility relates to the cross-linguistic suffixing preference that has long been noted (Greenberg, 1957), perhaps due to processing or learning biases (Cutler et al., 1985; Hawkins & Gilligan, 1988; Martin & Culbertson, 2020; Ramscar, 2013; St Clair et al., 2009). But whatever the underlying origin, such structural asymmetries in the lexicon—the lexicon's “informational bias”—may explain why readers identify words more accurately when fixating closer to word onset (the optimal viewing position; Brysbaert & Nazir, 2005; Hyönä & Bertram, 2011; O'Regan et al., 1984) and why readers prefer to target words closer to word onset during continuous reading (the preferred landing position; Ducrot & Pynte, 2002; McConkie et al., 1988; Vitu et al., 1990).

Nevertheless, evidence for a direct causal relationship between typical information spread and reading behavior is difficult to obtain for two principal reasons. First, during normal reading, a reader's decision about where to fixate a word may be influenced by several competing factors, including prior sentence context (Balota et al., 1985), preview of the upcoming word from the parafovea (Hyönä et al., 1989; Schotter et al., 2011; Underwood et al., 1990), saccade planning (Engbert & Krügel, 2010; Krügel & Engbert, 2014; McDonald & Shillcock, 2004), and constraints on perception (Bouma, 1973; McConkie & Rayner,

* Corresponding author at: Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom.

E-mail address: jon.carr@rhul.ac.uk (J.W. Carr).

1975). Of particular relevance is the finding that humans have an asymmetric visual span (typically a right-visual-field advantage), which has primarily been attributed to two possible (and non-mutually exclusive) origins: cerebral lateralization making it easier to process linguistic content in the right visual field (Brybaert & D'Ydewalle, 1988; Bub & Lewine, 1988; Ellis, 2004; Van der Haegen et al., 2013) and/or learned reading habits relating to reading direction (Huey, 1900; Mishkin & Forgays, 1952; Pollatsek et al., 1981). We take a neutral stance on this debate, referring to any such asymmetry as the reader's "perceptual bias," which we construe broadly to include anything that the human reader brings to the table in the process of visual word recognition. Importantly, this perceptual bias of the reader—like the informational bias of the lexicon, as described above—also typically predicts a preference for left-of-center fixation, making the two biases difficult to disentangle.

The second reason why it is difficult to test for a causal connection between the structure of the lexicon and reading behavior is that doing so necessitates the comparison of two or more natural languages. However, direct comparison of any two languages will be fraught with challenges because they will differ in many more ways than information distribution alone. For example, to date, the best available evidence has come from comparisons of languages written in left-to-right vs. right-to-left scripts (Deutsch & Rayner, 1999; Farid & Grainger, 1996; Pollatsek et al., 1981). In a right-to-left script (e.g., Arabic or Hebrew), the informative linguistic content is typically located on the right, predicting a preference for right-of-center fixation, albeit modulated to some extent by the aforementioned perceptual factors (Brybaert & Nazir, 2005; Nazir et al., 2004). However, comparison of languages written in different directions does not hold all factors constant. For example, the Arabic and Hebrew scripts are both abjads (scripts in which vowel sounds are represented by diacritics or, more commonly, not at all), which dramatically changes how much information is carried by each letter. Arabic is a cursive script, which may promote word decomposition at a higher level of analysis than individual letters (Farid & Grainger, 1996). Hebrew has a complex, non-concatenative morphology, which may result in different mechanisms of word recognition (Velan & Frost, 2011). Readers of Arabic and Hebrew are also likely to have at least somewhat regular exposure to left-to-right reading due to the widespread use of the Latin script, especially among the types of people who typically take part in psychological experiments (Ducrot & Pynte, 2002; Siéroff & Haehnel-Benoliel, 2015).

These two sets of issues—the complex interplay of several factors that determine where a reader will fixate a word during ordinary reading, along with the many confounds inherent to comparing natural languages—have made it difficult to evaluate the longstanding claim that lexical constraint plays a role in fixation patterns, as hypothesized by O'Regan (1981, p. 298). Herein, we attempt to evaluate this claim by taking a rather different approach. Rather than compare reading behavior in natural languages, we construct two artificial lexicons—two idealized test cases—that are identical except for the way in which information is spread laterally across words, eliminating any possible linguistic confounds (e.g., differences in orthography, script direction, morphosyntactic structure, etc.). To support our analyses, we set out a formal model of visual word recognition that explicitly captures the perceptual bias (i.e., whatever the reader brings to the table) and informational bias (i.e., whatever the language brings to the table), allowing us to study the independent effects of perception and information, as well as their interaction. Ultimately, we show that the informational structure of the lexicon has a causal effect on how people identify and target words, and that this effect can be explained through a general computational principle of uncertainty minimization.

Cognitive model and cross-linguistic analysis

We began by formulating a high-level Bayesian cognitive model of visual word recognition that explicitly encodes the perceptual and

informational biases. This model allowed us to explore how the two biases interact and understand how this interaction is likely to play out in natural languages. Furthermore, we used the model to aid in the interpretation of Experiment 1 and to help us make concrete predictions about Experiment 2.

Method

The model shares similarities with several other "Bayesian reader" models (Bicknell & Levy, 2012; Norris, 2006, 2009; Norris & Kinoshita, 2012; Smith et al., 2010; Valdois et al., 2021). The model conceptualizes visual word recognition as a simple two-stage process, comprised of perception and inference (see Fig. 1A). In the perception stage, the reader attempts to identify the letters in the word, forming a noisy percept. In the inference stage, the reader combines this perceptual evidence with prior information about what the word might plausibly be, forming a posterior over the lexicon from which it can choose a good candidate word.

Model description

The reader's lexicon W consists of n words of some constant length m , which are written using the set of symbols S . During perception, the constituent letters of a target word t may be corrupted according to the function Φ , which we refer to as the reader's "perceptual filter." This function dictates the probability of correctly perceiving the letter in position i given a fixation in position j , and is defined as follows:

$$\Phi(i|j) = \begin{cases} \left(\alpha - \frac{1}{|S|}\right) \exp[\beta(\gamma - 1)(i - j)] + \frac{1}{|S|} & \text{if } i > j \\ \left(\alpha - \frac{1}{|S|}\right) \exp[\beta(\gamma + 1)(i - j)] + \frac{1}{|S|} & \text{otherwise.} \end{cases} \quad (1)$$

The use of the exponential function here results in the probability of successfully identifying a character gradually approaching chance with distance. The perceptual filter, which is illustrated in Fig. 1B, is controlled by three parameters:

- $\alpha \in \left[\frac{1}{|S|}, 1\right)$ controls the probability that the reader will correctly identify the character under fixation within some fixed period of time (e.g., 50 ms).
- $\beta > 0$ controls the rate at which this probability approaches chance ($\frac{1}{|S|}$) with distance from the fixation position. Larger values result in a more rapid drop in probability and therefore a narrower visual span.
- $\gamma \in (-1, 1)$ controls the symmetry of the visual span. When $\gamma = 0$ the probability of correct letter identification drops symmetrically to the left and right (i.e., no perceptual bias). When γ is positive, the reader is more likely to correctly perceive letters to the right of fixation compared to letters to the left (i.e., a right-visual-field advantage). The reverse is true when γ is negative.

If a character is perceived incorrectly (i.e., with probability $1 - \Phi(i|j)$), one of the other $|S| - 1$ symbols will be perceived instead with uniform probability, $(1 - \Phi(i|j)) / (|S| - 1)$. This represents a simple model of perception that could be extended by, for example, taking letter confusability into account.

At the inference stage, the reader combines its noisy perceptual evidence with prior information about word frequency, forming a posterior over the lexicon. Given a percept p and knowledge of the fixation position j , the posterior probability of a hypothesized word w is given by Bayes' rule: $\Pr(w|p, j) \propto \Pr(p|w, j)\Pr(w)$. The likelihood term, $\Pr(p|w, j)$, denotes the probability that the perceptual filter would have given rise to percept p given that the true target was w fixated in position j . Following the definition of perception outlined above, the likelihood is therefore given by

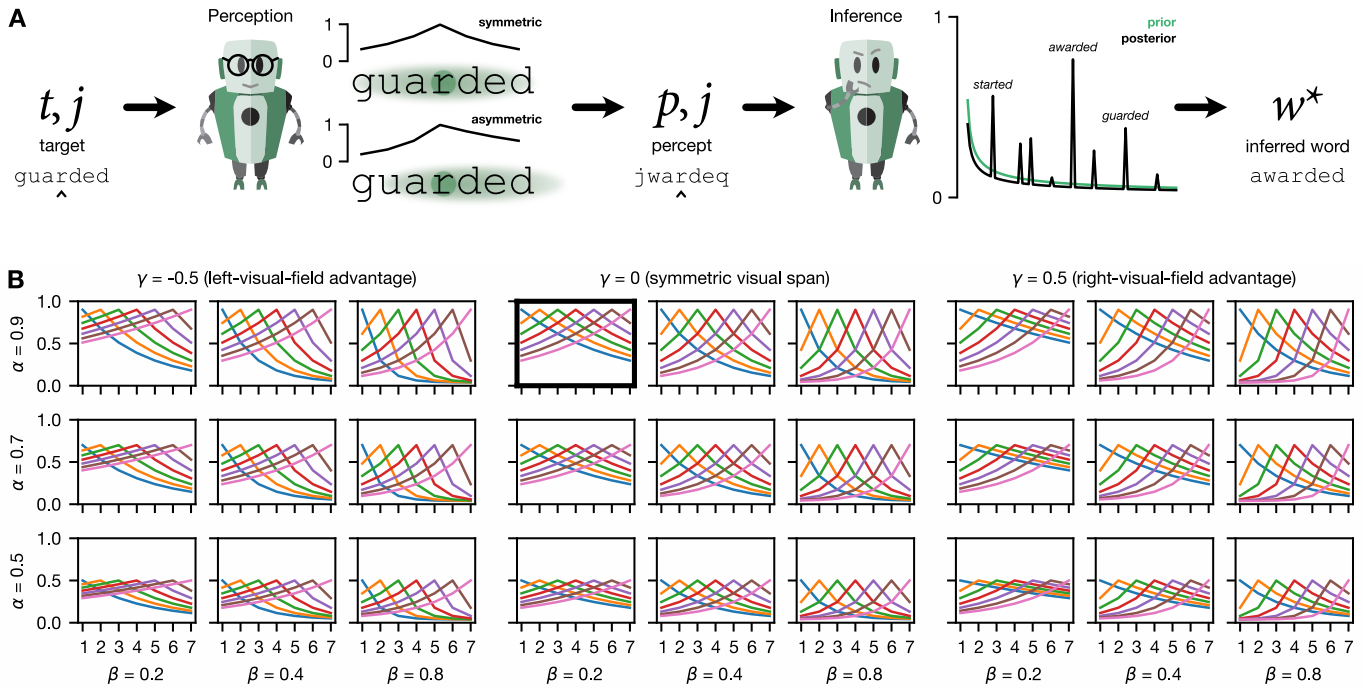


Fig. 1. **A** Illustration of a single trial in the core modeling framework. A target word t is fixated in position j and passes through the reader’s perceptual filter, resulting in percept p . Given this percept, the reader makes an inference about the target word by updating its frequency-based prior in light of the perceptual evidence. In the example, the reader is exposed to *guarded* in central position but incorrectly infers the word *awarded*. **B** Each panel represents the perceptual filter $\Phi(i|j)$ under a particular setting of α , β , and γ and shows the probability (Φ ; y-axes) of correctly identifying the character in positions 1 through 7 (i ; x-axes) given a fixation in positions 1 through 7 (j ; blue through pink). For example, if the parameters are set to $\alpha = 0.9$, $\beta = 0.2$, and $\gamma = 0$ (the panel highlighted with a thicker frame) and the reader is fixating in central position ($j = 4$; red curve), then the probability of correctly perceiving the characters in initial ($i = 1$) or final ($i = 7$) position is around 0.5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\Pr(p|w, j) = \prod_{i=1}^m \begin{cases} \Phi(i|j) & \text{if } p(i) = w(i) \\ \frac{1 - \Phi(i|j)}{|S| - 1} & \text{otherwise,} \end{cases} \quad (2)$$

where $p(i)$ and $w(i)$ represent the i th characters of p and w . The prior term, $\Pr(w)$, denotes the probability of a hypothesized word being true before the observation of any perceptual evidence and is set to the relative frequency of the word:

$$\Pr(w) = \frac{F(w)}{\sum_{w \in W} F(w)}. \quad (3)$$

This represents a simple assumption about the prior information available to the reader in the case of isolated word reading, but the model could be extended here by incorporating prior information from sentence context or parafoveal preview. Finally, the model reader infers word $w^* = \arg \max_{w \in W} \Pr(w|p, j)$, the word with maximum posterior probability.

Measure of uncertainty

During a single instance of visual word recognition, the uncertainty experienced by the reader may be quantified as the entropy of its posterior:

$$H(W|p, j) = - \sum_{w \in W} \Pr(w|p, j) \log[\Pr(w|p, j)]. \quad (4)$$

If the reader’s posterior tends to concentrate much of the probability mass on a single word, then uncertainty will be low; in contrast, if the reader’s posterior distributes the probability mass fairly across a variety of words, uncertainty will be high. The expectation of this quantity—that is, averaging over all words the reader might encounter and all percepts that might be formed in response to those words—may then be defined as follows:

$$U(j) = \sum_{t \in W} \sum_{p \in P} \Pr(t) \Pr(p|t, j) H(W|p, j). \quad (5)$$

In other words, we define the level of uncertainty a reader can expect to experience when fixating in position j as average posterior entropy,¹ weighted by the probability of a particular target word occurring ($\Pr(t)$; Equation 3) and the probability of a particular percept being formed given that target ($\Pr(p|t, j)$; Equation 2). The letter position that, on average, minimizes uncertainty is thus given by $\arg \min_{j=1}^m U(j)$. Our contention is that human readers develop implicit knowledge of the position that minimizes uncertainty (i.e., taking into account both the statistical structure of the lexicon and the reader’s own perceptual bias) and that they use this knowledge to maximize reading efficiency. See [Alhama et al. \(2019\)](#) for another approach to measuring uncertainty in this context.

Cross-linguistic datasets

To test the model on natural language data, we obtained word lists and frequency information for Dutch, English, German, Greek, Hebrew, Italian, Polish, and Spanish from word frequency databases based on TV and movie subtitles ([Brysaert et al., 2011](#); [Brysaert & New, 2009](#); [Crepaldi et al., 2015](#); [Cuetos et al., 2011](#); [Dimitropoulou et al., 2010](#); [Keuleers et al., 2010](#); [Mandera et al., 2014](#); [van Paridon & Thompson, 2021](#)). In addition, we reasoned that a prefixing language would be a good candidate for right-heavy structure; however, the only prefixing language for which we could obtain corpus data was Swahili ([Hurskainen, 2016](#)), which has been classified as a “weakly prefixing” language

¹ Since Equation 5 is intractable for even modest symbol sets and word lengths (i.e., because the number of possible percepts $|P| = |S|^m$), we estimate its value by simulating some large number of reading events.

(Dryer, 2013). We applied a number of preprocessing steps to the raw data from these studies, which included stripping case and accents from the words and filtering out words that used non-native characters (e.g., the letter κ in Italian). We then extracted the most frequent 5- to 9-letter words in each language. To ensure that the lexicon sizes would be consistent and comparable (despite the differing sizes of the datasets), we extracted the most frequent 3000 words for each word length in each language.²

Results

To check our basic assumptions, we instantiated the model with the English seven-letter lexicon and ran 100,000 simulated reading events of the words *guarded* and *concern* in initial ($j = 4$), central ($j = 4$), and final ($j = 7$) positions. Table 1 lists the top ten most common word inferences that the reader made. Under a symmetric visual span ($\alpha = 0.9, \beta = 0.2, \gamma = 0$), the reader is most likely to make the correct inference when fixating centrally because this tends to give the best view of the word. Accuracy in initial and final positions tends to be lower but is, crucially, dependent on the information structure of the word. For example, *guarded* is relatively informative on the left, so accuracy is higher given an initial fixation (57.9%) compared to a final fixation (23.3%). The opposite is true of *concern*: An initial fixation results in 53.5% accuracy, while a final fixation results in 72.8% accuracy.

This effect is further modulated by the reader's perceptual bias. Under a right-visual-field advantage ($\alpha = 0.9, \beta = 0.2, \gamma = 0.5$), accuracy is improved by shifting the fixation point further left to take full advantage of the bias (i.e., to bring the full word into the visual span). The result is that, in the case of left-heavy words, such as *guarded*, accuracy improves in initial position, but becomes very poor in final position. Interestingly, however, the word *concern* does not suffer in the same way: Accuracy is improved in initial position (because the right-visual-field advantage makes it possible to capture the crucial characters at the end of the word), but it also remains relatively high in final position (because this is where the most informative content is). This is an important dynamic that will also play out in our experiments.

The results presented in Table 1 are for just two words. We now scale up and consider the lexicon as a whole. Fig. 2 shows uncertainty by fixation position for the seven-letter words in four of the languages, English, Swahili, Polish, and Hebrew (see Fig. S1 in the Supplementary Material for all languages and word lengths). The uncertainty curves presented in black ($\gamma = 0$; no perceptual bias) take on a characteristic \cup -shape because, in general, fixating the center yields a better overall view than fixating the edges. Further, we find that these \cup -shapes may be asymmetric—and more so in some languages than others. In English, the uncertainty curve is left-heavy because left-heavy words are more prevalent than right-heavy ones. Swahili has a flatter distribution of information, Polish is slightly right-heavy, and Hebrew is very right-heavy.³ When the model reader is instantiated with a right-visual-field advantage ($\gamma = 0.5$; green curves), it becomes more advantageous to fixate further left, even in the case of Hebrew where the right-heavy distribution of information is overwhelmed.⁴ Of course, our setting of γ to 0.5 is merely illustrative at this point, but we will estimate its value from experimental data in the next section.

These results suggest two things. Firstly, there is at least some cross-linguistic variation in information distribution (see also Shafir et al.,

² One limitation of our model is that each word length must be considered separately.

³ Note that, in Hebrew, the right corresponds to the start of the words. In addition, vowel letters are typically omitted in writing, resulting in higher overall levels of uncertainty.

⁴ That being said, an account of the visual span based purely on reading habits might posit a left-visual-field advantage in the case of Hebrew, so for reference we also show results for $\gamma = -0.5$ in the Supplementary Material.

2022), which would therefore predict cross-linguistic variation in word targeting behavior. Secondly, any effect of information structure can easily be overshadowed by a sufficiently strong asymmetry in the visual span. As such, we set out to test the prediction that information distribution changes reading behavior by constructing two artificial lexicons—two ideal test cases—in which we can carefully control how information is distributed, while holding all other factors constant.

Experiment 1

The primary goal of our first experiment was to establish whether it is possible to experimentally manipulate the optimal viewing position effect with artificial lexicons. In addition, we wanted to estimate appropriate values for the perceptual parameters of the cognitive model, allowing us to make more concrete predictions about where participants ought to fixate if they seek to minimize uncertainty (predictions that we test in Experiment 2).

Method

Experiment 1 was not preregistered, since our goals were proof of concept, prediction generation, and parameter estimation.

Participants

Sixty participants were recruited via the Prolific platform and were paid £3.00 for participation (equivalent to a rate of £7.50 per hour based on the median completion time of 24 min). In addition to this base rate, participants could receive up to £1.20 in additional bonuses as detailed below (median bonus: £0.99). The most common first languages were: Portuguese (30%), English (15%), Polish (8.3%), Spanish (6.7%), and Greek (5%). Only two participants reported knowledge of a language that is written in a right-to-left script. The experiment was approved by the SISSA Ethics Committee (protocol number: 10027; date: 26/04/2021) and was conducted in accordance with all relevant ethical regulations. All participants provided informed consent.

Stimuli

We designed a pair of artificial lexicons—one left-heavy and one right-heavy—that would yield distinct, testable predictions in terms of where uncertainty will be minimized. Each lexicon consists of eight seven-letter words, the underlying structure of which is shown in Table 2. Each word followed a CCVCVCC pattern (e.g., *SVIMUKS*) that always began and ended with the letter *S*. These outer *S*'s can be thought of as lateral flankers that carry no information, eliminating potential issues relating to the special status of initial and final letters (Johnson & Eisler, 2012; White et al., 2008) or perceptual crowding effects (Pelli & Tillman, 2008). The five internal letters were constructed from a set of nine consonant letters, $\{B, D, G, K, M, N, P, T, V\}$, and a set of six vowel letters, $\{A, E, I, O, U, Y\}$. These two sets were permuted for each participant, such that the surface forms (examples of which are shown in Table 2) were different for each participant, although they always had the same underlying information structure. The characters that may occupy each position were chosen so that information is distributed asymmetrically. For example, across the eight words in the left-heavy lexicon, each of the letters in position 2 occurs twice; thus, this letter position conveys $-\log_2 \frac{2}{8} = 2$ bits of information about word identity. In position 6, by comparison, the letter c_6 occurs in all eight words, so this letter position conveys $-\log_2 \frac{8}{8} = 0$ bits of information. Importantly, the two lexicons are simply mirror-images of each other, and are thus identical except for the lateral distribution of information.

In order to present the stimuli at a consistent size across devices, participants were asked to place a credit card on the screen and adjust an on-screen image so that it matched the size of the physical card (Li et al., 2020). The words were presented in upper-case Courier New with a physical width of approximately 10 mm per letter. Participants were

Table 1
Top ten word inferences when fixating guarded and concern in initial, central, and final positions.

Symmetric visual span ($\gamma = 0$)											
Target word <i>t: guarded</i>						Target word <i>t: concern</i>					
Initial ($j = 4$)		Central ($j = 4$)		Final ($j = 7$)		Initial ($j = 4$)		Central ($j = 4$)		Final ($j = 7$)	
57.9%	guarded	64.7%	guarded	23.3%	guarded	53.5%	concern	80.8%	concern	72.8%	concern
4.8%	grandma	7.3%	started	14.1%	started	6.3%	control	1.6%	concert	2.2%	between
3.6%	guessed	3.2%	quarter	11.6%	married	3.4%	country	1.1%	conceal	1.5%	chicken
3.0%	getting	2.8%	married	11.5%	decided	2.5%	college	0.9%	concede	1.3%	popcorn
2.6%	glasses	1.7%	learned	2.8%	learned	1.6%	contact	0.8%	chicken	1.3%	pattern
1.8%	gunshot	1.5%	awarded	1.9%	hundred	1.5%	concert	0.7%	dancers	1.0%	goddamn
1.6%	granted	0.9%	charles	1.9%	wounded	1.2%	chicken	0.7%	concept	0.9%	captain
1.2%	grabbed	0.8%	charged	1.5%	changed	1.2%	confess	0.6%	sincere	0.9%	western
1.0%	quarter	0.8%	boarded	1.3%	crowded	1.2%	concept	0.6%	vincent	0.8%	shouldn
0.9%	goodbye	0.8%	grandma	1.3%	husband	1.1%	contest	0.5%	dancing	0.6%	lincoln

Right-visual-field advantage ($\gamma = 0.5$)											
Target word <i>t: guarded</i>						Target word <i>t: concern</i>					
Initial ($j = 4$)		Central ($j = 4$)		Final ($j = 7$)		Initial ($j = 4$)		Central ($j = 4$)		Final ($j = 7$)	
79.3%	guarded	65.8%	guarded	16.5%	decided	77.7%	concern	85.4%	concern	58.5%	concern
2.0%	guessed	6.8%	started	15.4%	married	1.8%	control	1.1%	concert	5.4%	between
1.5%	grandma	3.2%	married	12.8%	started	1.4%	concert	0.7%	sincere	2.4%	pattern
1.3%	quarter	2.2%	awarded	8.7%	guarded	1.2%	country	0.7%	dancers	2.1%	captain
1.1%	granted	2.1%	boarded	3.0%	learned	1.1%	conceal	0.6%	chicken	2.1%	chicken
1.0%	glasses	1.8%	quarter	2.5%	wounded	1.0%	college	0.5%	conceal	2.1%	shouldn
0.9%	started	1.8%	learned	2.4%	hundred	0.8%	concede	0.5%	popcorn	1.7%	goddamn
0.8%	grabbed	1.2%	decided	2.4%	changed	0.7%	concept	0.4%	concede	1.7%	popcorn
0.5%	gunshot	1.1%	charged	2.1%	husband	0.6%	condemn	0.4%	lincoln	1.6%	western
0.4%	learned	0.7%	charles	1.4%	worried	0.6%	contest	0.3%	vincent	1.0%	kitchen

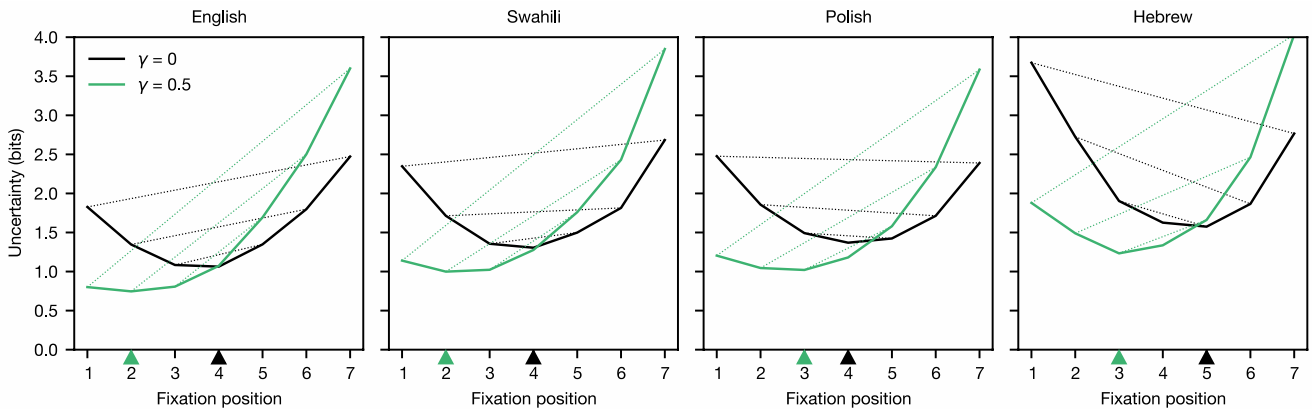


Fig. 2. Uncertainty by fixation position for the seven-letter words in English, Swahili, Polish, and Hebrew. The triangles on the x-axis mark the position of minimum uncertainty. Under a symmetric visual span (black; $\gamma = 0$), uncertainty is generally minimized in central position. Under a right-visual-field advantage (green; $\gamma = 0.5$), the position of minimum uncertainty is shifted further left. N.B., in all cases, including Hebrew, the x-axis represents the physical left-to-right character position. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Artificial lexicon structures with examples of possible surface forms (Experiments 1 and 2).

Left-heavy lexicon						Right-heavy lexicon									
Underlying structure			Example			Underlying structure			Example						
S	c_1	v_1	c_5	v_5	c_9	S	SNYBEVS	S	c_9	v_5	c_5	v_1	c_1	S	SVEBYNS
S	c_2	v_2	c_5	v_5	c_9	S	STOBEVS	S	c_9	v_5	c_5	v_2	c_2	S	SVEBOTS
S	c_3	v_3	c_6	v_5	c_9	S	SGUPEVS	S	c_9	v_5	c_6	v_3	c_3	S	SVEPUGS
S	c_4	v_4	c_6	v_5	c_9	S	SKAPEVS	S	c_9	v_5	c_6	v_4	c_4	S	SVEPAKS
S	c_3	v_1	c_7	v_6	c_9	S	SGYDIVS	S	c_9	v_6	c_7	v_1	c_3	S	SVIDYGS
S	c_1	v_2	c_7	v_6	c_9	S	SNODIVS	S	c_9	v_6	c_7	v_2	c_1	S	SVIDONS
S	c_4	v_3	c_8	v_6	c_9	S	SKUMIVS	S	c_9	v_6	c_8	v_3	c_4	S	SVIMUKS
S	c_2	v_4	c_8	v_6	c_9	S	STAMIVS	S	c_9	v_6	c_8	v_4	c_2	S	SVIMATS

asked to sit approximately 50 cm from the display, resulting in each letter occupying approximately 1.15 degrees of visual angle. The corresponding object stimuli were taken from the Novel Object and Unusual Name database (Horst & Hout, 2016) and were presented in grayscale at a size of approximately 40 × 40 mm.

Procedure

Participants were told that they would be learning an “alien” language and that their goal was to learn the words for a set of eight alien objects. In the training phase, participants were taught one of the two lexicons (Fig. 3). During training, the participant was passively exposed to four word–object pairs in quick succession. Each passive exposure trial lasted 2000 ms with the word appearing after a 500 ms delay so that the participant’s gaze was first drawn to the object and then to the word (i.e., the word itself was shown for 1500 ms). The participant was then shown one of the eight words (not necessarily from the previous four passive exposure trials) and had to click on the appropriate object, which we refer to as a “mini-test.” The purpose of these mini-test trials was to keep participants actively engaged in the learning process and to allow us to track learning performance over the course of the training period. Regardless of whether or not the participant chose the correct answer, the picture for the correct object remained on screen for one second, while the remaining object pictures disappeared, providing feedback and reinforcing the correct mapping. The participant was awarded a bonus of £0.01 for a correct answer. This procedure was repeated 64 times, such that each of the word–object pairs was passively exposed 32 times (256 passive exposure trials in total) and tested eight times (64 mini-test trials in total).

Participants were then tested on their ability to identify the words in each of the seven character positions (Fig. 4). In each test trial, the participant was first asked to look at a fixation point positioned horizontally center. After a randomly determined delay of between 1000 and 3000 ms, one of the eight words was displayed on screen for 50 ms and the participant was asked to select the corresponding object for that word, just as in the previous mini-test trials. Since 50 ms is below the time required to plan and execute a saccade (Rayner, 1998), this forces participants to identify the words based on a single fixation in an experimenter-controlled position. Feedback and bonusing was also provided as before. Crucially, the word was presented with some degree of eccentricity: One of the seven letters was aligned with the position of the fixation point, such that the word itself may appear shifted to the left or right. Each of the seven characters from each of the eight words was tested once in fixation position, resulting in 56 test trials.

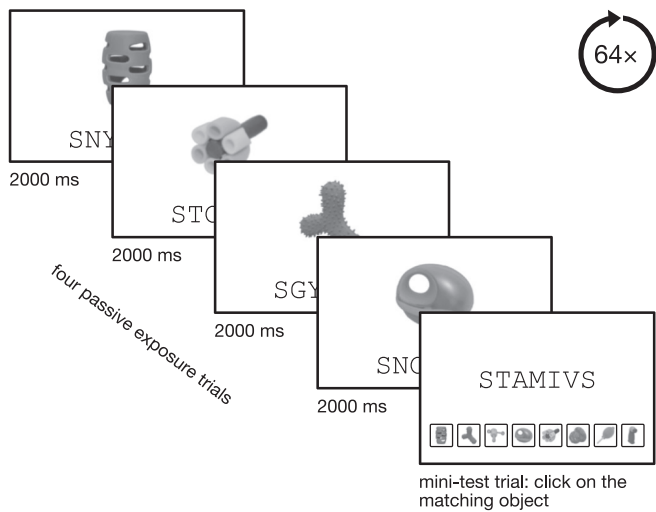


Fig. 3. Training regimen (all experiments). Participants complete a mixture of passive exposure trials and mini-test trials in which they have to select the appropriate object for a word.

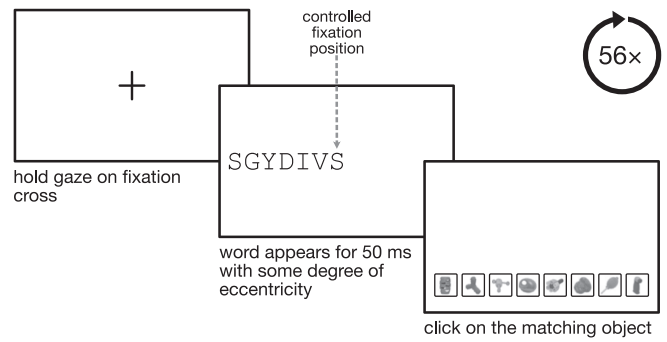


Fig. 4. Controlled fixation test trial (Experiment 1). The participant fixates a fixation point and, after a short delay, a word is displayed for 50 ms. The word is positioned such that one of the seven characters is aligned with the fixation point. In this example, the word *SGYDIVS* is fixated in the final, uninformative position, making it hard to identify.

Results

The majority of participants (51/60) made zero or one error during the final block of training, indicating good learning of the object–word pairs. The remaining nine participants, who made two or more errors, were excluded from further analysis. Our rationale with this exclusion criterion was that we can only meaningfully measure a participant’s ability to identify words in the test phase if they acquired the object–word pairs to a high degree of proficiency in the training phase (i.e., because correct object selection is only an index of correct word recognition). After these exclusions, the dataset that entered into our analysis included 1232 trials across 22 participants in the left-heavy condition and 1624 trials across 29 participants in the right-heavy condition.

Overall test accuracy was slightly lower in the left-heavy condition (mean: 0.8, 95% HDI: 0.75–0.84) compared to the right-heavy condition (mean: 0.85, 95% HDI: 0.81–0.89), which is to be expected under a right-visual-field advantage.⁵ Fig. 5 breaks down test accuracy by

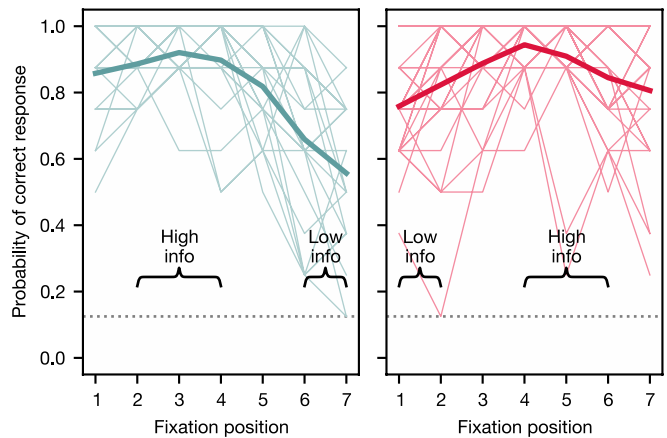


Fig. 5. Mean accuracy by fixation position for each condition in Experiment 1. The thin, light-colored lines show results for individual participants. In the left-heavy condition (blue), participants are most accurate when fixating left-of-center where the words are more informative. In the right-heavy condition (red), participants are most accurate in central position, despite the fact that the words are more informative right-of-center. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

⁵ Recall from our model simulations that, under a right-visual-field advantage, identification is easier on right-heavy words.

fixation position. In the case of the left-heavy condition, accuracy was highest in position 3 and lowest in position 7. The overall shape of this accuracy curve may be explained by perception, information, or both. The left-of-center area offers a good viewing position of the word as a whole (i.e., assuming a right-visual-field advantage), but the left-of-center area is also information rich. Likewise, the final position, where accuracy was lowest, offers a poor view of the word and contains low information content. In the case of the right-heavy condition, where we might naively expect to see mirrored results, we found that accuracy was actually maximized in central position and remained comparatively high even in position 1, where the information content is lowest. This may be explained by an interaction between perception and information: Accuracy is high in central position because this offers a good overall view of the word, including the informative content to the right; accuracy remains high in initial position because, assuming a right-visual-field advantage, the informative content near the end of the word can still be accurately captured.

In other words, in the left-heavy condition perception and information are aligned and reinforce each other, making a left-of-center fixation most beneficial. In the right-heavy condition, by contrast, the two factors cancel out: The structure of the lexicon makes it easier to identify words right-of-center, but the right-visual-field advantage makes a left-of-center fixation more advantageous. The net effect is that central fixation is most beneficial. To provide a quantitative confirmation of this explanation, we now fit the parameters of our cognitive model to the experimental data.

To perform the model fit, we introduced a noise parameter, $\epsilon \in (0, 1)$, in addition to the three perceptual parameters described earlier (α , β , and γ). This parameter represents the probability of participants making selection errors (e.g., mapping the inferred word to the wrong meaning or accidentally clicking the wrong button) and helps to improve our estimates of the perceptual parameters by accounting for other sources of noise separately. Given the assumptions of the model \mathcal{M} and some setting of the four model parameters, collectively denoted θ , the likelihood of observing experimental dataset D is given by

$$\Pr(D|\mathcal{M}, \theta) = \prod_{\langle t,j \rangle} \sum_{w \in \mathcal{W}} \begin{cases} \Pr(w'|t,j)(1 - \epsilon) & \text{if } w' = w \\ \Pr(w'|t,j) \frac{\epsilon}{n - 1} & \text{otherwise,} \end{cases} \quad (6)$$

where $\langle t, j, w \rangle$ represents each trial in D (the target word, the position in which that word was fixated, and the participant's inference as indicated by the object they clicked on). The term $\Pr(w'|t, j)$ specifies the probability that the model reader would infer w' when given the same input as the participant (the target word t fixated in position j).⁶ Therefore, the summation in Equation 6 accounts for all ways in which the model reader could make the same inference as the participant on a given trial – either by inferring the same word as the participant and sticking to that choice with probability $1 - \epsilon$ or by inferring some other word and accidentally switching to the same choice as the participant with probability $\epsilon/(n - 1)$.

To place a prior on each parameter, we chose a beta distribution (transformed to the relevant parameter bounds) that was representative of the information available to us before running the experiment. The priors are illustrated by the dashed gray lines in Fig. 6 and were motivated as follows:

- $\alpha \sim \text{Beta}(8, 2)$: We expected readers to have a fairly high probability of correctly identifying the character under fixation, probably well beyond 60%.
- $\beta \sim \text{Beta}(2, 8)$: We expected this parameter to be on the low end of the scale, corresponding to visual span that is several letters wide.

⁶ Formally this is given by $\sum_{p \in P} \Pr(p|t, j) \Pr(w'|p, j)$. However, in practice we estimate its value by simulating 10,000 reading events.

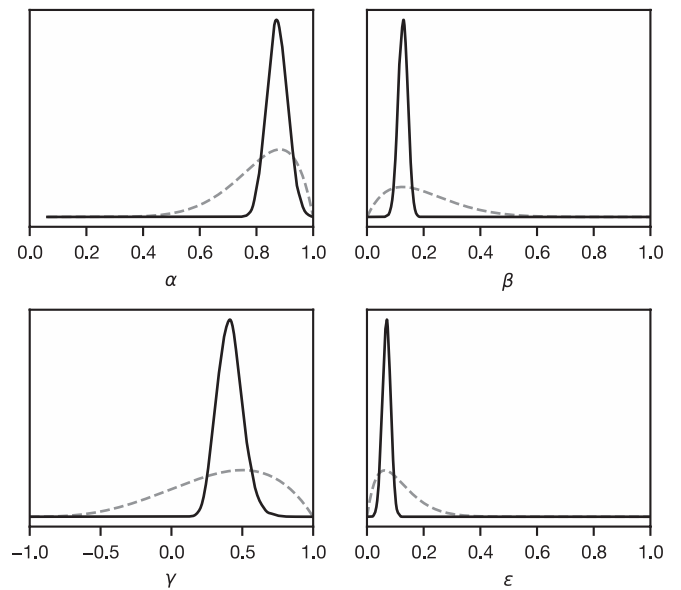


Fig. 6. Prior (dashed) and posterior (solid) densities for each model parameter. The positive estimate of γ confirms that participants have a right-visual-field advantage and quantifies its strength.

Small values (<0.01) and large values (>0.5) are implausible because these would correspond to overly wide and overly narrow visual spans respectively.

- $\gamma \sim \text{Beta}(4, 2)$: Following the prior literature, we expected readers to be better at identifying characters to the right of fixation, resulting in a positive γ value. However, since it was unclear how strongly this model parameter should be set, we opted for a broad prior that peaks at 0.5. We also wanted to remain relatively open-minded to the possibility of a symmetric visual span (i.e., $\gamma = 0$); perhaps, for example, the artificial language learning paradigm would be insufficient to reveal a visual span asymmetry.
- $\epsilon \sim \text{Beta}(2, 16)$: We expected selection errors to be fairly rare (e.g., less than 20%), especially since we had already excluded participants who demonstrated poor learning during the training phase.

The joint posterior distribution over the parameter space, $\Pr(\theta|\mathcal{M}, D)$, was estimated using the Python package PyMC with six chains of 2500 samples (sequential sampling, all ESS > 1000 , all $\hat{R} = 1$) and is plotted in Fig. 6. Crucially, γ was estimated to have a positive value of 0.41 (95% HDI: 0.25–0.59). Not only does this mean that our experimental dataset is best explained by assuming a right-visual-field advantage, as expected from the prior literature, but it also quantifies the strength of this advantage in the context of our experimental setup and modeling framework. Complete posterior parameter estimates are given in Table S1 in the Supplementary Material.

We also performed the model fit with uniform priors and obtained almost identical results. As an additional sanity check, we estimated the posterior using the data from each experimental condition independently, that is $\Pr(\theta|\mathcal{M}, D_{\text{left}})$ and $\Pr(\theta|\mathcal{M}, D_{\text{right}})$. These posteriors showed a high degree of overlap, suggesting that they reflect a single underlying perceptual filter that is common to all participants regardless of the lexicon they were exposed to in training (see Fig. S2 in the Supplementary Material).

Forward simulations of the experiment using parameter values drawn from the posterior can be seen in Fig. 7. These simulations are able to retrodict overall accuracy levels fairly well and capture the distinctive shapes of the two conditions. Importantly, the experimental results fall neatly within the posterior predictive distribution, meaning that the pattern of results we observed is exactly what one would expect to see when the right-visual-field advantage interacts with the

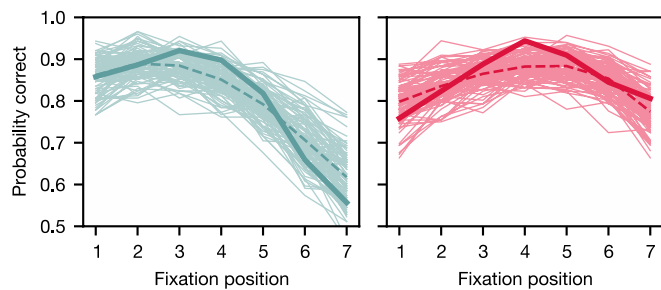


Fig. 7. Posterior predictive checks of the model fit. Each thin line depicts the mean accuracy curve from a simulated run of the experiment using parameter values drawn from the posterior. The dashed line shows the mean of 100 simulated runs; the thick line shows the actual experimental findings.

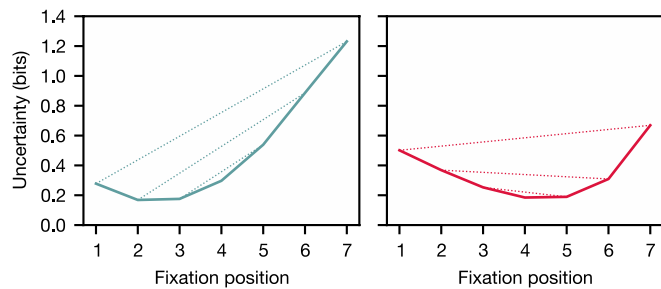


Fig. 8. Reader uncertainty by fixation position. These curves predict where uncertainty will be minimized after accounting for both the structure of the artificial lexicons and the best fitting parameters of the perceptual filter.

information distributions of the two constructed lexicons.

Finally, Fig. 8 shows predicted uncertainty by fixation position for the left- and right-heavy lexicons, incorporating the perceptual asymmetry estimated from the Experiment 1 data. The shapes of these uncertainty curves are distinctive in two important ways, yielding testable predictions. Firstly, participants exposed to the left-heavy lexicon should prefer to fixate around positions 2 or 3, since this is where uncertainty is minimized, while participants exposed to the right-heavy lexicon should prefer to fixate more centrally in positions 4 or 5. Secondly, and less obviously, the uncertainty curve for the right-heavy lexicon is comparatively flat, which may result in lower pressure to land exactly on the position of minimum uncertainty (i.e., because uncertainty is relatively low regardless of fixation position). This may lead to more dispersed landing positions in the right-heavy condition.

Experiment 2

Our first experiment showed that the classic optimal viewing position effect (e.g., Brysbaert et al., 1996) can be replicated with artificial languages: Accuracy in word recognition is a function of how much information is accessible at a given position and the shape of the visual span. But are readers able to learn which position minimizes uncertainty across the lexicon as a whole, and do they actively target this position? To answer these questions, we conducted a second experiment with the same stimuli and training procedure but a new type of test phase in which participants could, like Experiment 1, view the words for at most 50 ms but, unlike Experiment 1, freely target them in any position. Based on the predictions derived from our first experiment, we made two hypotheses:

1. Participants exposed to the right-heavy lexicon will target the words further right.
2. Participants exposed to the right-heavy lexicon will have more dispersed landing positions.

Method

The experiment was preregistered at https://aspredicted.org/BS2_6N2; we made no deviations from the preregistered analysis plan.

Participants

Eighty participants were recruited via our local participant pool and were paid €10 for participation (equivalent to a rate of €12 per hour based on the median completion time of 50 min). Participants were predominantly native speakers of Italian. The experiment was approved by the SISSA Ethics Committee (protocol number: 28135; date: 24/12/2021) and was conducted in accordance with all relevant ethical regulations. All participants provided informed consent.

Procedure

The stimuli and training phase were identical to Experiment 1, but participants completed a different type of test phase, as illustrated in Fig. 9. To initiate a trial, the participant had to hold their gaze for 2000 ms within an 18 px radius of a fixation point at the center of the screen. A target word then appeared either above or below the fixation point in a vertically random (but horizontally centered) position, and the word disappeared within 50 ms of its bounding box first being entered. The participant’s task was to locate and identify the word, and then select the corresponding alien object from the full array of objects, just as in Experiment 1. Feedback was provided as before. Each participant completed 64 test trials (eight trials for each of the eight words).

During the test phase, participants’ eye movements were recorded using an EyeLink 1000 Plus eye tracker (SR Research, Toronto, Canada) with monocular (dominant eye) recording at a 1000 Hz sampling rate. Participants used a headrest positioned 57 cm from the screen, and the words were presented at a width of 36 px per character (approximately 11 mm or 1.1 degrees of visual angle). To maintain a high level of precision and provide participants with frequent opportunities to relax and adjust their head position, the eye tracker was recalibrated at least every eight trials (or more often as required) using 13-point calibration. The participant’s initial landing position on a word was defined as the pixel distance between the first fixation recorded inside the word’s bounding box and the left edge of that bounding box.

Statistical model

To evaluate the two hypotheses, we constructed the following multilevel Bayesian statistical model:

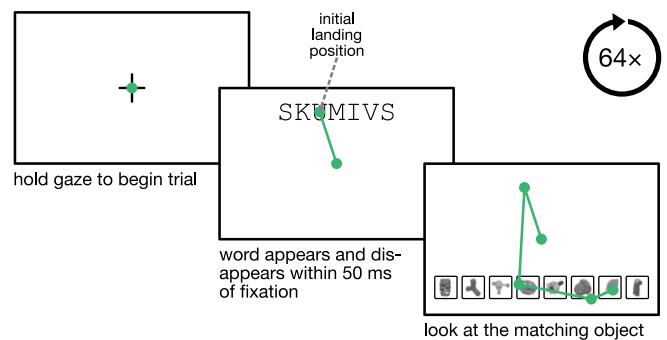


Fig. 9. Free fixation test trial (Experiments 2 and 3). The participant fixates a fixation point until a word flashes up in a random (but horizontally centered) position above or below the fixation point. The word disappears within 50 ms of being fixated, such that the participant must identify the word using a single fixation. The participant’s gaze path is shown in green. In this example, the participant targets the word SKUMIVS in third position. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\begin{aligned}
x_i &\sim \text{Normal}(\mu_j, \sigma_j) \\
\mu_j &\sim \text{Normal}(\tau_k, \zeta) \\
\sigma_j &\sim \text{Gamma}(\delta_k, \xi) \\
\tau_{\text{left}} &\sim \text{Normal}(72, 20) \\
\tau_{\text{right}} &\sim \text{Normal}(144, 20) \\
\zeta &\sim \text{Exponential}(0.1) \\
\delta_{\text{left}} &\sim \text{Gamma}(20, 8) \\
\delta_{\text{right}} &\sim \text{Gamma}(30, 8) \\
\xi &\sim \text{Exponential}(0.1) \\
\Delta(\tau) &= \tau_{\text{right}} - \tau_{\text{left}} \\
\Delta(\delta) &= \delta_{\text{right}} - \delta_{\text{left}}
\end{aligned}$$

where i indexes trials by participant j , and j indexes participants assigned to condition k . The model states that the pixel landing positions x_i are normally distributed according to a mean μ_j and standard deviation σ_j that is specific to each participant. These by-participant μ 's and σ 's are also drawn from normal (or normal-like) distributions, representing across-participant variation in targeting behavior.⁷ The average target pixel position is represented by the hyperparameter τ and the average dispersion is represented by the hyperparameter δ , both stratified by condition, while ζ and ξ place regularizing hyperpriors on the variation in across-participant behavior. The deterministic parameters $\Delta(\tau)$ and $\Delta(\delta)$ represent the difference in targeting and dispersion behavior between the two conditions. Formally, Hypothesis 1 states that $\Delta(\tau) > 0$ and Hypothesis 2 states that $\Delta(\delta) > 0$.

For the purpose of making discrete inferential decisions about the hypotheses, we followed the ROPE (region of practical equivalence) and HDI (highest density interval) approach (Kruschke, 2015). We set the ROPEs based on a minimally interesting effect size (one quarter of the width of a character). The ROPEs on $\Delta(\tau)$ and $\Delta(\delta)$ were $[-9, 9]$ and $[-4, 4]$ respectively. Our decision rules were as follows: If the 95% HDI lies entirely above the ROPE, we accept the hypothesis; if it lies entirely within the ROPE, we reject the hypothesis; and if it overlaps the ROPE bounds, we consider the evidence indeterminate.

Our sample size was determined by a precision-based sequential sampling procedure, as recommended in Kruschke (2015). We planned to stop data collection once the widths of the 95% HDIs on $\Delta(\tau)$ and $\Delta(\delta)$ shrank below 10 px and 5 px respectively, which we deemed to be a desirable level of precision. Due to resource limitations, however, we capped the number of participants at 80. This precision-based stopping criterion has no bearing on accepting or rejecting the hypotheses; it simply allows us to stop the experiment early if a high level of precision is obtained, regardless of whether the inferential decision is positive, negative, or indeterminate.

Results

We did not meet the level of precision required to terminate the experiment early and therefore collected data from a total of 80 participants.⁸ Performance during the training phase was broadly the same as Experiment 1, and we applied the same training-based exclusion criterion resulting in seven participants being excluded. In the case of five participants, it was necessary to terminate the experimental session early due to calibration difficulties with the eye tracker (nevertheless, the trials that we did collect from these participants still enter into our analyses). Additionally, in a small number of trials (typically three or four trials per participant), no initial landing position could be identified because the participant's gaze swept over the word (causing it to disappear) without actually forming a fixation within the word

boundary and within the word presentation time. The final dataset included 2119 trials across 37 participants in the left-heavy condition and 2165 trials across 36 participants in the right-heavy condition.

Test accuracy was close to ceiling and almost identical between the left (mean: 0.97, 95% HDI: 0.95–0.98) and right (mean: 0.98, 95% HDI: 0.97–0.99) conditions. Higher accuracy levels were expected in Experiment 2 (compared to Experiment 1) because participants should now be targeting the words in their ideal locations, resulting in accurate word recognition on almost every trial. Fig. 10 and Fig. 11 plot the distribution of initial landing positions by condition and are immediately suggestive of a difference in reading behavior.

The statistical model was fit using the Python package PyMC with six chains of 2500 samples (NUTS, all ESS > 10000, all $\hat{R} = 1$). The results are plotted in Fig. 12. In terms of Hypothesis 1, there is a clear effect of condition with the posterior difference, $\Delta(\tau)$, estimated at 34 px (95% HDI: 27–41) or approximately one character position. Participants exposed to the left-heavy lexicon tended to land 94 px into the word (95% HDI: 89–99; character position 3), while participants exposed to the right-heavy lexicon tended to land 128 px into the word (95% HDI: 123–133; character position 4). In terms of Hypothesis 2, the HDI overlaps the ROPE bounds so we consider the evidence to be indeterminate: On the one hand, 98.5% of the posterior probability mass lies above 0, indicating that dispersion is greater in the right-heavy condition as hypothesized; but, on the other hand, 90.1% of the posterior probability mass falls within the bounds of the ROPE, suggesting that, if there is an effect, it is likely to be of a negligible magnitude. Specifically, in the left-heavy condition, dispersion was 21.7 px (95% HDI: 20.1–23.3), and in the right-heavy condition, dispersion was 24.2 px (95% HDI: 22.6–25.9), yielding a posterior difference of 2.5 px (95% HDI: 0.3–4.8). Complete posterior parameter estimates are given in Table S2 in the Supplementary Material.

To check that the posteriors were not unduly influenced by our choice of priors, we also fit the model using uniform priors on τ and δ and obtained almost identical results (see Fig. S3 in the Supplementary Material). We also found no meaningful difference in the results if the across-participant variation parameters (ζ and ξ) were stratified by condition (see Fig. S4 in the Supplementary Material).

Overall, our results show that the information distribution of the learned language has a causal effect on how participants target words. Interestingly, participants do not simply target the part of the word that contains the most information; rather, they target the position that will minimize overall uncertainty, taking into account both the lexicon's information spread as well as the reader's own perceptual constraints, as

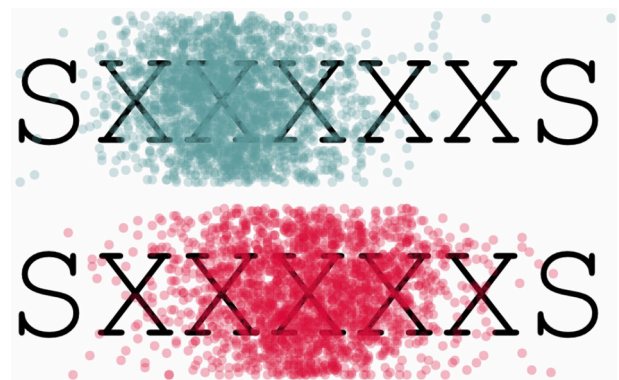


Fig. 10. All initial landing positions for the left-heavy (blue) and right-heavy (red) conditions. All words began and ended with the letter s, but the internal letters were different for each word/participant and are therefore represented here as x's. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

⁷ Note that we follow the convention of parameterizing the Gammas with a mean and standard deviation.

⁸ The 95% HDI widths on $\Delta(\tau)$ and $\Delta(\delta)$ were 13.56 px and 4.51 px respectively, so we met the second criterion and were close to meeting the first. Note also that, although we did not reach the desired level of precision to stop data collection early, the posteriors are still fully interpretable.

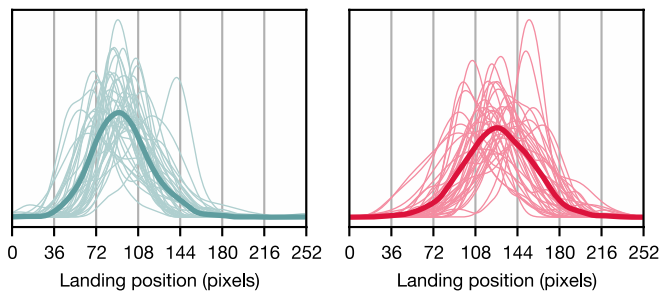


Fig. 11. Kernel density plots of initial landing position. The thin, light-colored curves represent individual participants. Participants in the left-heavy condition (blue) tend to land left-of-center, while participants in the right-heavy condition (red) tend to land centrally. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predicted by our model. Note, for example, that participants exposed to the right-heavy lexicon primarily target the center of the word, presumably because this position still permits good access to the most information-heavy part of the word due to the right-visual-field advantage.

Experiment 3

One limitation of our artificial lexicons is that they consist purely of left-heavy or right-heavy words, whereas natural language lexicons include a mixture of both kinds of words. English, for example, has left-heavy words like *guarded* and right-heavy words like *concern*, although left-heavy words are more common leading to the overall bias in the English lexicon’s information distribution. Our third experiment addresses this issue by adopting two new artificial lexicons in which the majority (but not all) of the words are left- or right-heavy. Like Experiment 2, we hypothesized that participants exposed to the right-dominant lexicon would target the words further right than those exposed to the left-dominant lexicon. We did not make a formal

hypothesis about dispersion, since the results of Experiment 2 were suggestive of a weak or null effect on dispersion.

Method

Except for the new lexicons, the methods were identical to that of Experiment 2. The experiment was preregistered at https://aspredicted.org/6RQ_GM7; we made no deviations from the preregistered analysis plan.

Participants

Fifty-four participants were recruited via our local participant pool and were paid €10 for participation (equivalent to a rate of €12.50 per hour based on the median completion time of 48 min). Participants were predominantly native speakers of Italian. The experiment was approved by the SISSA Ethics Committee (protocol number: 28135; date: 24/12/2021) and was conducted in accordance with all relevant ethical regulations. All participants provided informed consent.

Stimuli

Table 3 shows the structure of the two new lexicons. In each case, the first six words (i.e., the majority of words, 75%) use the dominant information distribution, while the final two words use the nondominant distribution. For example, in the left-dominant lexicon, the first six words can be thought of as containing a suffix (e.g., *-IFS*) and are therefore more informative on the left, while the final two words can be thought of as containing a prefix (e.g., *STI-*) and are therefore more informative on the right. Although the stimuli follow the same overall CCVCVCC pattern as before, it was necessary to introduce two additional consonant letters (*F* and *L*) and reduce the information contained in the central character, which is now fully uninformative.

Statistical model

Experiment 3 was analyzed using the following Bayesian statistical model:

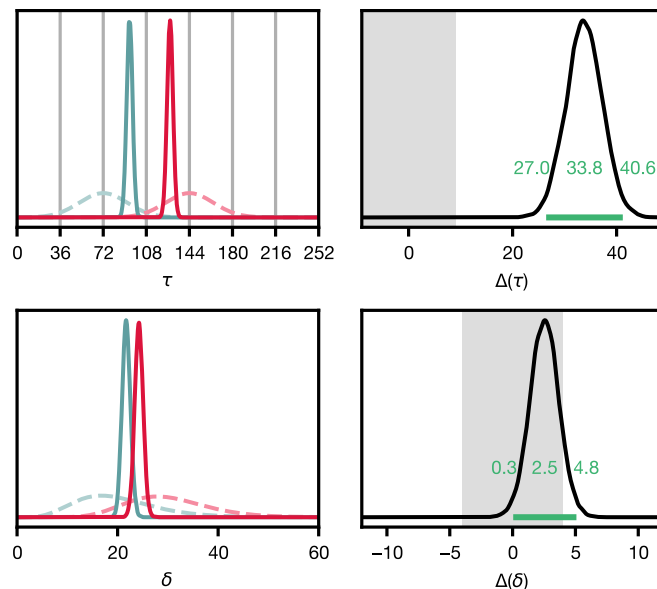


Fig. 12. Prior (dashed) and posterior (solid) densities for τ and δ , along with the posterior differences. The gray shaded area shows the ROPE and the green bar shows the 95% HDI. Participants in the left-heavy condition (blue) target position 3 and participants in the right-heavy condition (red) target position 4. The posterior probability mass for $\Delta(\tau)$ lies well above 0, indicating a clear difference between conditions. Dispersion is slightly larger in the right-heavy condition; however, since the 95% HDI overlaps the ROPE bounds, the evidence is considered indeterminate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Artificial lexicon structures with examples of possible surface forms (Experiment 3).

Left-dominant lexicon							Right-dominant lexicon								
Underlying structure						Example	Underlying structure						Example		
S	c ₁	v ₁	c ₈	v ₄	c ₉	S	SMOGIFS	S	c ₉	v ₄	c ₈	v ₁	c ₁	S	SFIGOMS
S	c ₂	v ₂	c ₈	v ₄	c ₉	S	SDUGIFS	S	c ₉	v ₄	c ₈	v ₂	c ₂	S	SFIGUDS
S	c ₃	v ₃	c ₈	v ₄	c ₉	S	SKAGIFS	S	c ₉	v ₄	c ₈	v ₃	c ₃	S	SFIGAKS
S	c ₄	v ₁	c ₈	v ₄	c ₉	S	SNOGIFS	S	c ₉	v ₄	c ₈	v ₁	c ₄	S	SFIGONS
S	c ₅	v ₂	c ₈	v ₄	c ₉	S	SLUGIFS	S	c ₉	v ₄	c ₈	v ₂	c ₅	S	SFIGULS
S	c ₆	v ₃	c ₈	v ₄	c ₉	S	SPAGIFS	S	c ₉	v ₄	c ₈	v ₃	c ₆	S	SFIGAPS
S	c ₇	v ₄	c ₈	v ₅	c ₁₀	S	STIGYBS	S	c ₁₀	v ₅	c ₈	v ₄	c ₇	S	SBYGITS
S	c ₇	v ₄	c ₈	v ₆	c ₁₁	S	STIGEVS	S	c ₁₁	v ₆	c ₈	v ₄	c ₇	S	SVEGITS

- $x_i \sim \text{Normal}(\mu_j, \sigma_j)$
- $\mu_j \sim \text{Normal}(\tau_k, \zeta)$
- $\sigma_j \sim \text{Gamma}(\delta, \xi)$
- $\tau_{\text{left}} \sim \text{Normal}(94, 30)$
- $\tau_{\text{right}} \sim \text{Normal}(128, 30)$
- $\delta \sim \text{Gamma}(23, 10)$
- $\zeta \sim \text{Exponential}(0.1)$
- $\xi \sim \text{Exponential}(0.1)$
- $\Delta(\tau) = \tau_{\text{right}} - \tau_{\text{left}}$

The model diverges from that of Experiment 2 in two main ways. First, we dropped the dispersion contrast—the estimate for the δ parameter is now pooled across both conditions. Second, our priors on τ and δ were updated to reflect what we had learned from Experiment 2; that is, we expected participants to target pixel 94 in the left-dominant condition and pixel 128 in the right-dominant condition, with around 23 px of dispersion. We also made a few other minor tweaks based our experience with Experiment 2. We increased the standard deviations on the τ and δ priors, since we felt our previous priors had been overly confident; we raised the precision criterion to 16 px, since we felt our previous criterion had been unnecessarily demanding; and we adjusted the ROPE bounds to $[-10, 10]$ to bring our specification into alignment with [Kruschke's \(2015\)](#) recommendation that the precision criterion be 80% of the ROPE width.

Results

Training performance was broadly the same as Experiments 1 and 2, and we applied the same training-based exclusion criterion resulting in three participants being excluded. As in Experiment 2, no initial landing position could be identified in a small number of cases (typically three or four trials per participant) because the participant did not form a fixation within the word boundary and within the word presentation time. The final dataset included 1497 trials across 25 participants in the left-heavy condition and 1583 trials across 26 participants in the right-heavy condition.

[Fig. 13](#) and [Fig. 14](#) plot the distribution of initial landing positions by condition, and the statistical results are plotted in [Fig. 15](#). The results continue to show a clear effect of condition on targeting behavior with the posterior difference, $\Delta(\tau)$, estimated at 26 px (95% HDI: 18–34). Participants exposed to the left-heavy lexicon tended to land 97 px into the word (95% HDI: 92–103; character position 3), while participants exposed to the right-heavy lexicon tended to land 124 px into the word (95% HDI: 118–129; character position 4). Complete posterior parameter estimates are given in [Table S3](#) in the [Supplementary Material](#).

We also fit the model using uniform priors on τ and δ and obtained almost identical results (see [Fig. S5](#) in the [Supplementary Material](#)), and we found no meaningful change in the results after stratifying all parameters by condition (see [Fig. S6](#) in the [Supplementary Material](#)). Although we did not make a formal hypothesis about dispersion, we did note some evidence of a difference ($\Delta(\delta) = 4.6$ px; 95% HDI: 1.3–7.8), although the effect was still not strong enough to meet the decision criteria we set out in Experiment 2. Perhaps more interestingly, there

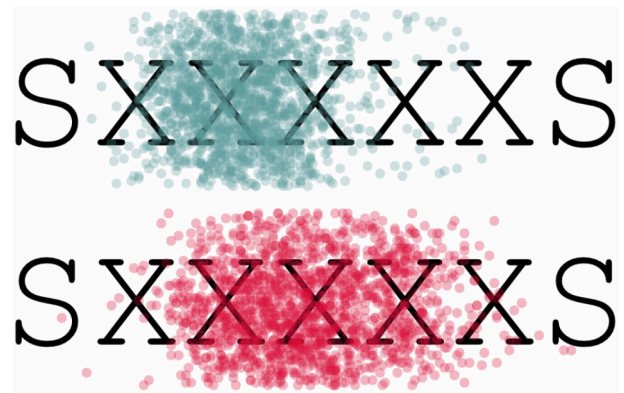


Fig. 13. All initial landing positions for the left-heavy (blue) and right-heavy (red) HDIs. All words began and ended with the letter s, but the internal letters were different for each word/participant and are therefore represented here as x's. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was some evidence to suggest that across-participant variation in targeting (ζ) was larger in the right-dominant condition ($\Delta(\zeta) = 6.8$ px; 95% HDI: 0.6–13.1), although, again, the evidence for this is not entirely conclusive. Nevertheless, it does appear in [Fig. 14](#) that participants in the right-dominant condition used a wider range of strategies than those in the left-dominant condition, with some targeting position 3, some targeting position 4, some targeting position 5, and some having a bimodal distribution targeting positions 3 and 5. It is not immediately clear why there might be a difference between conditions in this regard, but one possibility relates to carry-over from the participants' native language, Italian. If participants come into the study with an a-priori preference for targeting words left-of-center, it is perhaps more difficult for those assigned to the right-dominant condition to overcome this prior experience, especially since the minority-type words (i.e., the left-heavy words in the right-dominant lexicon) did indeed require a left-of-center fixation.

Overall, the results of Experiment 3 show that the effect of average information distribution on word targeting continues to hold even when the lexicon contains a mixture of both left- and right-heavy words. Participants exposed to the right-dominant lexicon, in which the majority of words are more informative on the right, fixated around 26 px further to the right than participants exposed to the left-dominant lexicon. This effect size was smaller than that observed in Experiment 2, but this may be explained by the fact that, under a mixed lexicon, participants need to hedge their bets, since they cannot know a priori whether or not the target word will conform to the dominant information distribution.

Discussion

Mounting evidence shows that language processing takes full advantage of probabilistic cues present in the language itself. One such

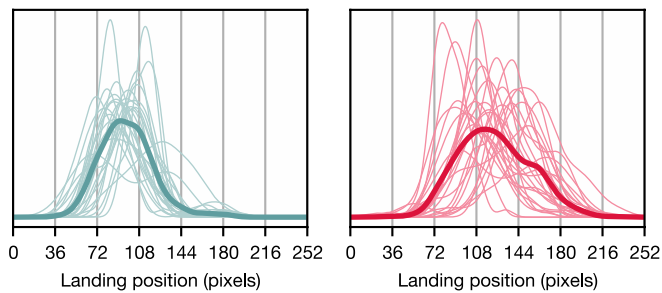


Fig. 14. Kernel density plots of initial landing position. The thin, light-colored curves represent individual participants. Participants in the left-heavy condition (blue) tend to land left-of-center, while participants in the right-heavy condition (red) tend to land centrally. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

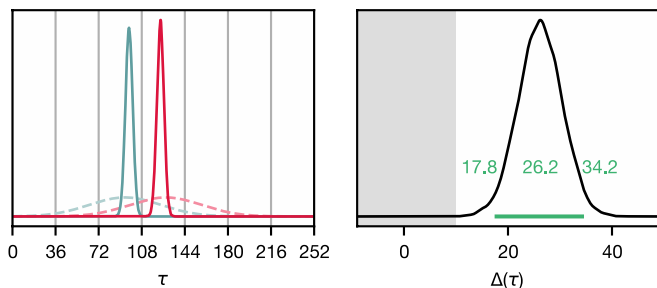


Fig. 15. Prior (dashed) and posterior (solid) densities for τ , along with the posterior difference $\Delta(\tau)$. The gray shaded area shows the ROPE and the green bar shows the 95% HDI. Participants in the left-heavy condition (blue) target position 3 and participants in the right-heavy condition (red) target position 4. The posterior probability mass for $\Delta(\tau)$ lies well above 0, indicating a clear difference between conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cue in the domain of reading is the way in which information about word identity is distributed across words, providing readers with expectations about which position tends to be most informative (i.e., minimizes uncertainty about word identity). Previous research has explored this issue by contrasting word processing in different languages, and while this has provided precious insight, the approach has struggled to disentangle information distribution from perceptual asymmetries and other important factors that distinguish human languages (e.g., morphology and reading direction). To circumvent this complexity, we adopted an artificial language learning approach, which allowed us to isolate the role of information distribution and test its *causal* effect on eye movement behavior.

We first devised a cognitive model of visual word recognition that predicts uncertainty as a function of the human perceptual bias and the linguistic informational bias. Applying this model to a selection of languages, we found that there is at least some cross-linguistic variation in how information is distributed within lexicons, which therefore predicts cross-linguistic variation in how readers ought to target words. We tested this prediction by submitting participants to one of two artificial lexicons, which differ only in how information is laterally distributed.

In Experiment 1, we focused on accuracy in word recognition. We demonstrated the feasibility of eliciting the classic optimal viewing position effect (Brybaert & Nazir, 2005; Hyönä & Bertram, 2011; O’Regan et al., 1984) using artificially constructed lexicons, and we showed that the effect could, in principle, be manipulated by changing the distribution of information. The interaction we observed between the right-visual-field advantage and the location of information in the words is in close alignment with similar experiments by Brybaert et al. (1996) using real languages (Dutch and French). Experiment 1 also

allowed us to estimate the parameters of the human visual span, as it pertains, at least, to our experimental paradigm and modeling framework. This revealed clear evidence in support of the right-visual-field advantage and also allowed us to calibrate our model in order to make concrete predictions about reading behavior under the assumption that readers aim to minimize uncertainty when targeting words.

In Experiment 2, we estimated the causal effect on reading behavior of being exposed to the right-heavy lexicon instead of the left-heavy lexicon. The effect was to shift participants’ initial landing positions around one character position to the right, which was broadly in line with our model predictions. Evidence for an effect on dispersion was unclear, although we note that either the presence or absence of a dispersion effect could have interesting theoretical ramifications. A difference in dispersion would suggest that participants are sensitive to how concentrated information is in one particular area, while no difference in dispersion would suggest that participants adhere to a maximizing rule—always target the position of minimum uncertainty, even if it is only marginally better than some other position. Future work could aim to resolve this.

In Experiment 3, we addressed one potential issue with our artificial lexicons, namely that the lexicons consisted purely of one type of word (either left-heavy or right-heavy), whereas natural language lexicons contain a mixture of words with different distributions of information. The results of this experiment continued to show a clear difference in landing position, suggesting that participants are able to pick up on a cue that is more statistical in nature. In nonpreregistered analyses, we found some evidence to suggest that there is more variation in participants’ targeting strategies in the right-dominant condition. This could be interpreted as a native-language carry-over effect. Participants in the left-dominant condition can simply stick to their normal patterns of behavior (target words left-of-center), while participants in the right-dominant condition must overcome their prior expectations. Individual differences in willingness to adapt to a new lexical environment might then explain the greater variation in targeting behavior. Future work could aim to investigate the issues of native-language carry-over and individual differences in this paradigm.

A broad distinction has been drawn in the literature between “fast-heuristic” accounts of targeting behavior, in which readers primarily target the center of an upcoming word, and “cognitive-processing” accounts, which incorporate ongoing processing of the upcoming word (Bicknell et al., 2020). Such ongoing processing may include, for example, predictions from prior sentence context (Balota et al., 1985) or information garnered through parafoveal preview (Hyönä et al., 1989; Schotter et al., 2011; Underwood et al., 1990). The present study represents a hybrid of these two accounts—a fast heuristic that is nevertheless rooted in experience with the language. Even in the absence of prior information about the identity of an upcoming word, readers can nevertheless target the position in which they expect to minimize uncertainty. This is not merely the position of maximum information content, nor the position that permits the best view of the word, but rather the position that optimizes the trade-off between the two.

Another interesting observation concerns the flexibility of the system. It is perhaps surprising that participants were able to overcome a lifetime of experience with just 15 min of exposure to these novel lexicons. However, related work has shown that the way participants attend to words can rapidly adapt to a new lexical environment (Ducrot & Pynte, 2002). Moreover, substantial cognitive flexibility has been observed in many other domains. For example, participants can overcome the cultural representation of time as extending from left to right with just five minutes of mirror-reversed reading (Casasanto & Bottini, 2014). These results point to human cognition as a flexible system that is able to respond quickly to a changing environment and account for volatile statistical cues.

The extent to which our findings may generalize to normal reading scenarios (as opposed to isolated word reading) remains an open question. In particular, the use of lexicon-level statistical cues is likely to be

supplemented by other sources of prior information, and these sources are likely to interact with each other in complex ways. The reader's prediction about the upcoming word based on prior sentence context might adjust how much weight they place on the more general language-level heuristic investigated here. If the upcoming word is very likely to be *concern*, it perhaps pays to discard what you know about the language in general, and instead target the word right-of-center where you can expect to gather maximum information. Likewise, if highly constraining information has already been gathered from the parafovea about the initial letters in the upcoming word, it might, again, pay to flout what is known about the language in general.

With these issues in mind, it will be important to verify our findings in more naturalistic settings where the myriad determinants of eye movement behaviors interact. One way to approach this is using large, cross-linguistic eye movement corpora such as MECO (Siegelman et al., 2022); indeed, work in this direction is already underway (Shafir et al., 2022). Alternatively, experimental paradigms, such as the one used here, could be further extended and refined to incorporate other pertinent factors. Our artificial words could be placed in predictive contexts, for example, and the availability of parafoveal information could be manipulated using the boundary paradigm. A particularly important issue will be to understand whether the oculomotor system is even precise enough to yield optimal landing positions based on the brain's prediction about where maximum information can be extracted. It has been shown for example, that a substantial proportion of saccades do not land on the targeted word, suggesting that the oculomotor system might introduce considerable noise in the targeting process (Engbert & Nuthmann, 2008).

The model we presented herein is conceptually similar to the Bayesian Reader (Norris, 2006, 2009; Norris & Kinoshita, 2012). Like that work, our model is firmly expressed at the computational level of description (Marr, 1982) and does not, therefore, make any strong commitments to particular neural processes or linguistic representations. Indeed, our model is even more simplified, dispensing with such features as positional uncertainty, reaction times, letter confusability, and crowding effects. The main feature our model adds—like other Bayesian models (Bicknell & Levy, 2012; Smith et al., 2010; Valdois et al., 2021)—is that the probability of correctly identifying a letter is proportional to its distance from fixation. These choices were all motivated by keeping the model tightly focused around our particular domain of interest, that is, the relative contributions of perception and information to visual word recognition and the causal effect of information structure on word targeting. A nice feature of our model is that it does not make strong assumptions about linguistic representation and relies entirely on the system's sensitivity to statistics in the lexicon. For example, the reader will be biased toward inferring *ed*, *er*, or *es* at the end of a word—especially if it has a bad view of the end of the word—simply by virtue of the fact that these are common word endings in the lexicon, and not because of any built-in *n*-gram chunking or awareness of morphology.

The present work shows that sophisticated linguistic processes, like word recognition and targeting, may be explained in terms of simple, general-purpose cognitive heuristics such as maximal efficiency in the collection of information. This is in line with other work in the domain of visual processing, showing that the brain allocates more resources to aspects of the input that are more informative (Hermundstad et al., 2014; Simoncelli & Olshausen, 2001). It also sits well with the observation that reading capitalizes on neural circuitry that was originally devoted, both ontogenetically and phylogenetically, to non-linguistic visual objects such as faces and tools (Dehaene & Cohen, 2011; Hervais-Adelman et al., 2019). Although this might seem obvious (printed words are visual objects after all), it has widespread implications for the way we conceive of the lexical system. Much recent work has captured linguistic phenomena in terms of sensitivity to the statistics of the input. For example, word learning is affected by letter co-occurrence statistics (Chetail, 2017; Lelonekiewicz et al., 2020), and

aspects of orthographic processing that were long thought to be specific to letters and words emerge with entirely non-linguistic visual stimuli (Vidal et al., 2021).

Readers are also sensitive to the consistency in the mapping between spelling, sound, and meaning (Marelli et al., 2015; Ulicheva et al., 2020), and signs of orthographic processing more generally have been observed in both naive (Rajalingham et al., 2018) and trained (Grainger et al., 2012) primates. Moreover, the efficient organization of lexical semantics has been shown to emerge in computational agents as they interact and develop a shared language (Chaabouni et al., 2021). This research suggests, at least in principle, that intelligent agents, gifted only with relatively simple—and certainly not language-specific—computational strategies, can develop behaviors and communication systems that share core features with human language.

This does not necessarily mean that the linguistic constructs we have used for decades (e.g., graphemes, morphemes, even words) are misplaced. The general-purpose computational machinery referenced above, combined with the input it receives, might certainly converge on representations and processes that are not dissimilar from classical linguistic constructs. It is also surely true that language and reading are so pervasive in human experience that some domain-specific phenomena may be expected to emerge. For example, position coding does seem to be different between letters and other visual objects (Duñabeitia et al., 2012). However, such domain specificity might come from intensive exposure rather than an intrinsic difference in how the brain processes language; from experience and efficient statistical machines, rather than specialized neural machinery.

CRediT authorship contribution statement

Jon W. Carr: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Monica Fantini:** Investigation. **Lorena Perrotti:** Investigation. **Davide Crepaldi:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data and code for the model, cross-linguistic analysis, and experiments are publicly accessible at <https://osf.io/s3gev/>.

Acknowledgments

This work was funded by an ERC (European Research Council) Starting Grant (no. 679010, STATLEARN) and a FARE (Framework per l'Attrazione e il Rafforzamento delle Eccellenze per la Ricerca in Italia) 2016 grant (no. R164XBHNYR, CROWDLEARN), both awarded to DC.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jml.2024.104530>.

References

- Alhama, R. G., Siegelman, N., Frost, R., & Armstrong, B. C. (2019). The role of information in visual word recognition: A perceptually-constrained connectionist account. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual conference of the Cognitive Science Society* (pp. 83–89). Cognitive Science Society.

- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17(3), 364–390. [https://doi.org/10.1016/0010-0285\(85\)90013-1](https://doi.org/10.1016/0010-0285(85)90013-1)
- Bicknell, K., & Levy, R. (2012). Word predictability and frequency effects in a rational model of reading. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society* (pp. 126–131). Cognitive Science Society.
- Bicknell, K., Levy, R., & Rayner, K. (2020). Ongoing cognitive processing influences precise eye-movement targets in reading. *Psychological Science*, 31(4), 351–362. <https://doi.org/10.1177/0956797620901766>
- Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, 13(4), 767–782. [https://doi.org/10.1016/0042-6989\(73\)90041-2](https://doi.org/10.1016/0042-6989(73)90041-2)
- Brysaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424. <https://doi.org/10.1027/1618-3169/a000123>
- Brysaert, M., & D'Ydewalle, G. (1988). Callosal transmission in reading. In G. Lüer, U. Lass, & J. Shallo-Hoffman (Eds.), *Eye movement research: Physiological and psychological aspects* (pp. 246–266). Hogrefe.
- Brysaert, M., & Nazir, T. (2005). Visual constraints in written word recognition: Evidence from the optimal viewing-position effect. *Journal of Research in Reading*, 28(3), 216–228. <https://doi.org/10.1111/j.1467-9817.2005.00266.x>
- Brysaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Brysaert, 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(3), 385–395. <https://doi.org/10.1037/0894-4105.10.3.385>
- Bub, D. N., & Lewine, J. (1988). Different modes of word recognition in the left and right visual fields. *Brain and Language*, 33(1), 161–188. [https://doi.org/10.1016/0093-934X\(88\)90060-0](https://doi.org/10.1016/0093-934X(88)90060-0)
- Casasanto, D., & Bottini, R. (2014). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General*, 143(2), 473–479. <https://doi.org/10.1037/a0033297>
- Chaabouni, R., Kharitonov, E., Dupoux, E., & Baroni, M. (2021). Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, 118(12), Article e2016569118. <https://doi.org/10.1073/pnas.2016569118>
- Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, 163, 103–120. <https://doi.org/10.1016/j.cognition.2017.02.015>
- Clark, J. J., & O'Regan, J. K. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research*, 39(4), 843–857. [https://doi.org/10.1016/S0042-6989\(98\)00203-X](https://doi.org/10.1016/S0042-6989(98)00203-X)
- Crepaldi, D., Amenta, S., Mander, P., Keuleers, E., & Brysaert, M. (2015). *SUBTLEX-IT: Subtitle-based word frequency estimates for Italian*. Poster presentation. Annual meeting of the Italian association for experimental psychology, Rovereto, Italy.
- Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysaert, M. (2011). *SUBTLEX-ESP: Spanish word frequencies based on film subtitles*. *Psicologica*, 32, 133–143.
- Cutler, A., Hawkins, J. A., & Gilligan, G. (1985). The suffixing preference: A processing explanation. *Linguistics*, 23(5), 723–758. <https://doi.org/10.1515/ling.1985.23.5.723>
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254–262. <https://doi.org/10.1016/j.tics.2011.04.003>
- Deutsch, A., & Rayner, K. (1999). Initial fixation location effects in reading Hebrew words. *Language and Cognitive Processes*, 14(4), 393–421. <https://doi.org/10.1080/016909699386284>
- Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: The case of Greek. *Frontiers in Psychology*, 1, 1–12. <https://doi.org/10.3389/fpsyg.2010.00218>
- Dryer, M. S. (2013). Prefixing vs. suffixing in inflectional morphology. In M. S. Dryer, & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.
- Ducrot, S., & Pynte, J. (2002). What determines the eyes' landing position in words? *Perception & Psychophysics*, 64(7), 1130–1144. <https://doi.org/10.3758/BF03194762>
- Duñabeitia, J. A., Dimitropoulou, M., Grainger, J., Hernández, J. A., & Carreiras, M. (2012). Differential sensitivity of letters, numbers, and symbols to character transpositions. *Journal of Cognitive Neuroscience*, 24(7), 1610–1624. https://doi.org/10.1162/jocn_a.00180
- Ellis, A. W. (2004). Length, formats, neighbours, hemispheres, and the processing of words presented laterally or at fixation. *Brain and Language*, 88(3), 355–366. [https://doi.org/10.1016/S0093-934X\(03\)00166-4](https://doi.org/10.1016/S0093-934X(03)00166-4)
- Engbert, R., & Krügel, A. (2010). Readers use Bayesian estimation for eye movement control. *Psychological Science*, 21(3), 366–371. <https://doi.org/10.1177/0956797610362060>
- Engbert, R., & Nuthmann, A. (2008). Self-consistent estimation of mislocated fixations during reading. *PLoS ONE*, 3(2), Article e1534. <https://doi.org/10.1371/journal.pone.0001534>
- Farid, M., & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language*, 53(3), 351–368. <https://doi.org/10.1006/brln.1996.0053>
- Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in baboons (*Papio papio*). *Science*, 336(6078), 245–248. <https://doi.org/10.1126/science.1218152>
- Greenberg, J. H. (1957). Order of affixing: A study in general linguistics. In J. H. Greenberg (Ed.), *Essays in linguistics* (pp. 86–94). University of Chicago Press.
- Hawkins, J. A., & Gilligan, G. (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua*, 74(2–3), 219–259. [https://doi.org/10.1016/0024-3841\(88\)90060-5](https://doi.org/10.1016/0024-3841(88)90060-5)
- Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., & Tkačik, G. (2014). Variance predicts salience in central sensory processing. *eLife*, 3, Article e03722. <https://doi.org/10.7554/eLife.03722>
- Hervais-Adelman, A., Kumar, U., Mishra, R. K., Tripathi, V. N., Guleria, A., Singh, J. P., Eisner, F., & Huettig, F. (2019). Learning to read recycles visual cortical networks without destruction. *Science Advances*, 5(9), eaax0262. <https://doi.org/10.1126/sciadv.aax0262>
- Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409. <https://doi.org/10.3758/s13428-015-0647-3>
- Huey, E. B. (1900). On the psychology and physiology of reading. I. *The American Journal of Psychology*, 11(3), 283–302. <https://doi.org/10.2307/1412745>
- Hurskainen, A. (2016). Helsinki Corpus of Swahili 2.0 Not Annotated Version. Kielipankki. <http://urn.fi/urn:nbn:fi:lb-2016011302>
- Hyönä, J., & Bertram, R. (2011). Optimal viewing position effects in reading Finnish. *Vision Research*, 51(11), 1279–1287. <https://doi.org/10.1016/j.visres.2011.04.004>
- Hyönä, J., Niemi, P., & Underwood, G. (1989). Reading long words embedded in sentences: Informative word halves affects eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 15(1), 142–152. <https://doi.org/10.1037/0096-1523.15.1.142>
- Johnson, R. L., & Eisler, M. E. (2012). The importance of the first and last letter in words during sentence reading. *Acta Psychologica*, 141(3), 336–351. <https://doi.org/10.1016/j.actpsy.2012.09.013>
- Keuleers, E., Brysaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650. <https://doi.org/10.3758/BRM.42.3.643>
- Krügel, A., & Engbert, R. (2014). A model of saccadic landing positions in reading under the influence of sensory noise. *Visual Cognition*, 22(3–4), 334–353. <https://doi.org/10.1080/13506285.2014.894166>
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.
- Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language*, 115, Article 104152. <https://doi.org/10.1016/j.jml.2020.104152>
- Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific Reports*, 10, Article 904. <https://doi.org/10.1038/s41598-019-57204-1>
- Mander, P., Keuleers, E., Wodniecka, Z., & Brysaert, M. (2014). Subtlex-pl: Subtitle-based word frequency estimates for Polish. *Behavior Research Methods*, 47(2), 471–483. <https://doi.org/10.3758/s13428-014-0489-4>
- Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of orthography-semantics consistency on word recognition. *The Quarterly Journal of Experimental Psychology*, 68(8), 1571–1583. <https://doi.org/10.1080/17470218.2014.959709>
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1–2), 71–102. [https://doi.org/10.1016/0010-0277\(87\)90005-9](https://doi.org/10.1016/0010-0277(87)90005-9)
- Martin, A., & Culbertson, J. (2020). Revisiting the suffixing preference: Native-language affixation patterns influence perception of sequences. *Psychological Science*, 31(9), 1107–1116. <https://doi.org/10.1177/0956797620931108>
- McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, 28(10), 1107–1118. [https://doi.org/10.1016/0042-6989\(88\)90137-X](https://doi.org/10.1016/0042-6989(88)90137-X)
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17(6), 578–586. <https://doi.org/10.3758/BF03203972>
- McDonald, S. A., & Shillcock, R. C. (2004). The potential contribution of preplanned refixations to the preferred viewing location. *Perception & Psychophysics*, 66(6), 1033–1044. <https://doi.org/10.3758/BF03194993>
- Mishkin, M., & Forgy, D. G. (1952). Word recognition as a function of retinal locus. *Journal of Experimental Psychology*, 43(1), 43–48. <https://doi.org/10.1037/h0061361>
- Nazir, T. A., Ben-Boutayab, N., Decoppet, N., Deutsch, A., & Frost, R. (2004). Reading habits, perceptual learning, and recognition of printed words. *Brain and Language*, 88(3), 294–311. [https://doi.org/10.1016/S0093-934X\(03\)00168-8](https://doi.org/10.1016/S0093-934X(03)00168-8)
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2), 327–357. <https://doi.org/10.1037/0033-295X.113.2.327>
- Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, 116(1), 207–219. <https://doi.org/10.1037/a0014259>
- 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. <https://doi.org/10.1037/a0028450>

- O'Regan, J. K. (1981). The convenient viewing position hypothesis. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements, cognition, and visual perception* (pp. 289–298). Routledge.
- 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. <https://doi.org/10.1037/0096-1523.10.2.250>
- Paridon, J. van, & Thompson, B. (2021). subs2vec: Word embeddings from subtitles in 55 languages. *Behavior Research Methods*, 53(2), 629–655. <https://doi.org/10.3758/s13428-020-01406-3>
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11(10), 1129–1135. <https://doi.org/10.1038/nn.2187>
- Pollatsek, A., Bolozky, S., Well, A. D., & Rayner, K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and Language*, 14(1), 174–180. [https://doi.org/10.1016/0093-934X\(81\)90073-0](https://doi.org/10.1016/0093-934X(81)90073-0)
- Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience*, 38(33), 7255–7269. <https://doi.org/10.1523/JNEUROSCI.0388-18.2018>
- Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, 46(4), 377–396. <https://doi.org/10.2298/PSI1304377R>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Schotter, E. R., Angele, B., & Rayner, K. (2011). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, 74(1), 5–35. <https://doi.org/10.3758/s13414-011-0219-2>
- Shafir, T. K., Siegelman, N., Frost, R., & Armstrong, B. C. (2022). Predicting fixation locations in 43 languages based on perceptual constraints and information theory. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the annual conference of the Cognitive Science Society* (p. 3966). Cognitive Science Society.
- Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H.-D., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Kwon, N., Lõo, K., & Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The multilingual eye-movement corpus (MECO). *Behavior Research Methods*, 54, 2843–2863. <https://doi.org/10.3758/s13428-021-01772-6>
- Siéoroff, E., & Haehnel-Benoliel, N. (2015). Environmental script affects lateral asymmetry of word recognition: A study of French-Hebrew bilinguals tested in Israel and in France. *Laterality: Asymmetries of Body, Brain and Cognition*, 20(4), 389–417. <https://doi.org/10.1080/1357650X.2014.988220>
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1), 1193–1216. <https://doi.org/10.1146/annurev.neuro.24.1.1193>
- Smith, N. J., Chan, W.-H., & Levy, R. (2010). Is perceptual acuity asymmetric in isolated word recognition? Evidence from an ideal-observer reverse-engineering approach. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 1483–1488). Cognitive Science Society.
- Snowling, M. J., Hulme, C., & Nation, K. (2022). *The science of reading: A handbook* (2nd ed.). Wiley Blackwell.
- St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, 33(7), 1317–1329. <https://doi.org/10.1111/j.1551-6709.2009.01065.x>
- Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition*, 195, Article 103810. <https://doi.org/10.1016/j.cognition.2018.09.013>
- Underwood, G., Clews, S., & Everatt, J. (1990). How do readers know where to look next? Local information distributions influence eye fixations. *The Quarterly Journal of Experimental Psychology Section A*, 42(1), 39–65. <https://doi.org/10.1080/14640749008401207>
- Valdois, S., Phénix, T., Fort, M., & Diard, J. (2021). Atypical viewing position effect in developmental dyslexia: A behavioural and modelling investigation. *Cognitive Neuropsychology*, 38(5), 319–335. <https://doi.org/10.1080/02643294.2021.2004107>
- Van der Haegen, L., Cai, Q., Stevens, M. A., & Brysbaert, M. (2013). Interhemispheric communication influences reading behavior. *Journal of Cognitive Neuroscience*, 25(9), 1442–1452. https://doi.org/10.1162/jocn_a_00412
- Velan, H., & Frost, R. (2011). Words with and without internal structure: What determines the nature of orthographic and morphological processing? *Cognition*, 118(2), 141–156. <https://doi.org/10.1016/j.cognition.2010.11.013>
- Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology*, 31(6), 1261–1267. <https://doi.org/10.1016/j.cub.2020.12.017>
- Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & Psychophysics*, 47(6), 583–600. <https://doi.org/10.3758/BF03203111>
- White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1261–1276. <https://doi.org/10.1037/0096-1523.34.5.1261>
- Yeatman, J. D., & White, A. L. (2021). Reading: The confluence of vision and language. *Annual Review of Vision Science*, 7(1), 487–517. <https://doi.org/10.1146/annurev-vision-093019-113509>