Title

Visual Tuning for Letters

Author

Winsler, Kurt

Publication Date

2024

Visual Tuning for Letters

By

KURT WINSLER
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Steven J. Luck, Chair

George R. Mangun

John M. Henderson

Committee in Charge

2024

Acknowledgements

This dissertation, like me, is an effect of an infinite set of causes. Chief among these are the people who have taught and loved me over the last 31 years or so. Whatever credit is due for this collection of words is owed to them. To my family, by blood or otherwise, who have given me this life and the freedom to explore it. To Phil and Kathy and my undergraduate lab who set me on this weird (but enjoyable) path to science with everything I needed. To Steve, our lab, and so many of members of the Psychology and Center for Mind and Brain community who've done everything to cultivate this brilliant and collaborative environment that has been my intellectual home. And to my friends and partners, past and present, who have taught me so much about how to live and enjoy this life. I feel so lucky to have you all and your support over all these years. With all the love I have, thank you.

Abstract

The visual system is tuned by its inputs. The behavior of reading offers a unique way to examine tuning for visual representations (letters) because readers have massive experience recognizing letters in a systematic context (reading). One aspect of reading is that letters are highly crowded within words, which severely limits their perceptibility. Study 1 demonstrates how the recognition of letters can survive higher degrees of crowding compared to less familiar stimuli like inverted letters or Gabor patches, suggesting letter-specific tuning mechanisms that mitigate effects of visual crowding. This advantage for upright letters was particularly true for spatial locations closer to the fovea and in the right visual field; the locations that are most relied on for reading left-to-right orthographies. Study 2 follows up on this finding and shows that this pattern of reduced crowding for letters was increased for observers who score higher on measures of reading experience, which is also consistent with the idea that reading experience tunes letter representations in specific ways to reduce crowding. Study 3 examines the neural processing of individual upright and inverted letters using a combination of univariate ERP analyses and ERP decoding analyses. It shows that early in visual processing (before 100 ms) ERP decoding for upright letter identity is greater than for inverted letters, suggesting very early tuning for upright letters in the visual system. Further, this early difference was specific to upright letters in the fovea and the right visual field but not for letters in the left visual field. Overall, these behavioral and electrophysiological results illustrate that reading experience shapes the representation of letters in fundamental ways that reflect the constraints that reading puts on the visual system.

Table of Contents

Introduction

The visual world is complex, and the information humans must extract from it is diverse. Accordingly, the visual system has to be incredibly efficient as well as deeply flexible. To accomplish this, the visual system has evolved a richly interconnected and recurrent hierarchical structure that is tuned by the statistics of the environment it develops in. The initial properties of this processing stream—such as its retinotopic organization—are largely determined by genes and nonspecific prenatal experiences, whereas high level object representations are learned through experience with the visual environment; individuals are not born with intrinsically coded representations of specific visual objects. The structure of these high level representations must be learned based on patterns experienced in the outside world, and this experience is in part determined by the behaviors individuals engage in. Thus, individuals inherit an initialized visual system that is efficient, yet flexible and expressive enough to learn to represent the complex and dynamic visual information specific to that individual's world. Although it is clear that the information encoded at the start and end this process is largely driven by evolution and experience respectively, the information encoded in intermediate steps is determined by some mixture of inherited structure and experience-based tuning. Understanding these processes is at the crux of perception, yet uncovering the role of experience-based tuning is difficult given that it unfolds over an individual's lifetime.

The behavior of reading is an interesting and useful case for studying effects of expertise on visual representation. It has only recently been widespread in humans (e.g., global literacy passed 50% in the 1960s; van Zanden, et al., 2014), too recent for biological evolution to exert significant pressure on brain function. Hence, brain areas used for reading are co-opted from existing structures through the process of learning to read (see, e.g., Dehaene & Cohen, 2007). Reading is a demanding task for the visual system, yet given its importance for individuals in literate societies, it is highly practiced, leading to extremely high degrees of expertise. The behavior of reading is complex, but also highly structured: readers must perform quick and precise eye-movements to efficiently sample sentences, recognize letters and bind them into words, and integrate this lexical information with phonological, semantic, and syntactic structures. The difficulty of this behavior likely drives significant tuning in perceptual systems, while its systematic nature provides researchers with predictions about the nature of this tuning.

Here, I present a series of studies that aim to characterize the nature of tuning for letters in the human visual system, informed by the constraints that reading—the behavior that necessitates this tuning—puts on the system. One difficulty that the visual system faces during reading is that letters are highly crowded within words, which limits their perceptibility (Pelli, 2007), and the effects of visual crowding become more severe at spatial locations further from the fovea (Bouma 1970). Reading left-to-right scripts (e.g., English) involves disambiguating sets of visually similar letters not only in the fovea, but also in the parafoveal right visual field where readers sample upcoming information to plan eye movements (Schotter, Angele, and Rayner, 2012), to guide recognition of currently fixated words (Veldre & Andrews, 2016), and in some cases to recognize upcoming words without fixating them (Rayner et el., 2011). This leads to the prediction that reading may tune letter representations such that letters can be recognized under higher degrees of crowding than other stimuli, and this may be particularly true for the spatial locations in the visual field that are used for reading. Study 1 tests these predictions by comparing crowding distances for letters and other less-familiar stimuli across six locations in the visual field. Study 2 further tests this hypothesis by examining whether observers who score higher on measures of reading experience have reduced crowding for letters in the same 6 spatial locations. Finally, Study 3 tests whether early visual tuning for letters can be observed by comparing the time course of ERP decoding for upright and inverted letters presented in the fovea as well as in the left and right visual hemifields.

Chapter 1: Stimulus differences in visual crowding

In a classic study that documented a fundamental limit on visual perception, Bouma (1970) demonstrated that a peripheral letter could be accurately identified when presented in isolation but not when "crowded" by nearby flanker letters. Figure 1 provides a demonstration. Importantly, Bouma also demonstrated that the distance at which flankers crowd a target scales linearly with the distance of that target from fixation (which has subsequently been termed "Bouma's law"; Pelli and Tillman, 2008). Visual crowding is now known to occur for many types of stimuli and is a major source of functional limitation in real-world vision. It reflects fundamental properties of the visual system and provides clues to the nature of object recognition and visual representation (Levi 2008; Whitney and Levi, 2011; Manassi and Whitney, 2018). Bouma's original study and much subsequent research (Pelli et al., 2007; Chung, 2007; He, Legge, and Yu, 2013) has been concerned with crowding as a limitation on reading, because letters are highly crowded within words. Indeed, readers use crowded information from upcoming letters and words that are as many as 15 character spaces away from fixation (Rayner 2009), and there is now considerable evidence that crowding is a key limiting factor in both typical reading (Risse, 2014; Frömer et al., 2015; He & Legge, 2017) and disorders of reading such as dyslexia (Joo et al., 2018).

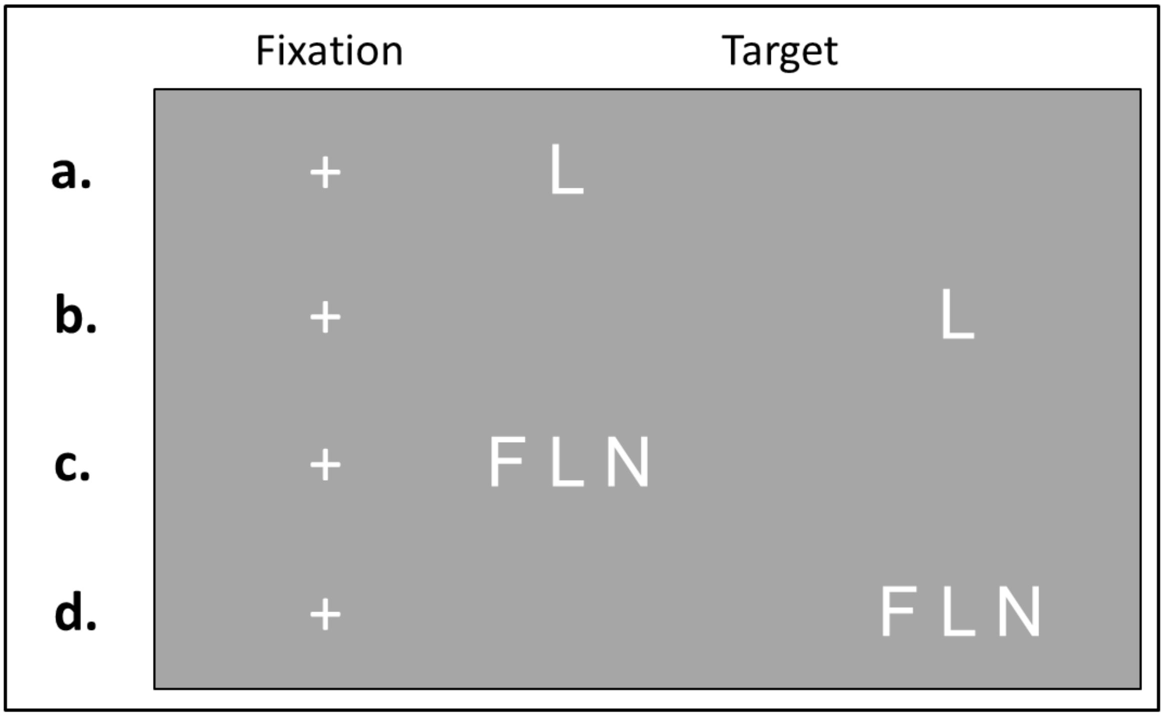

Figure 1. Demonstration of crowding. Keep your eyes on the central fixation crosshair for a given row while trying to perceive the letter(s) for that row. The letter L can be perceived quite well at both a near eccentricity (**a**) and a far eccentricity (**b**). However, the L becomes much more difficult to perceive when crowded by nearby flanker letters at the near eccentricity (**c**) and practically impossible to perceive at the far eccentricity (**d**). The fact that the L is easily perceived in isolation indicates that low-level visual acuity is not the reason for the difficulty of perceiving the L when it is crowded.

Here, we invert this relationship between crowding and reading to ask a different question: To what extent does our massive experience with reading affect crowding? It is clear that learning shapes visual processing at a high level (e.g., learning the forms of visual objects) as well as in a variety of lower level contexts (Sagi, 2011). However, the extent to which fundamental properties of vision such as crowding are influenced by experiences such as natural reading remains unknown. Reading offers a unique avenue to uncover the effects of experience on vision because it is a common part of daily life for many humans but was invented too recently to have driven significant brain evolution. In addition, reading is much more constrained than most common visual tasks. For example, proficient readers of the Latin script have viewed a relatively small set of letters millions of times, in a particular part of the visual field and attentional context (scanning across the horizontal meridian), and for a particular purpose (to bind letters into words, and ultimately into meaning). While this is no easy task for the untrained visual system, it becomes effortless for many people after years of practice. Developing this proficiency may involve systematic changes in fundamental aspects of vision such as crowding, adapting visual processing to the particular requirements of reading through neuroplasticity.

We focus on three aspects of reading that might lead to corresponding changes in crowding. First, reading involves massive experience with a relatively small number of visually-similar characters, which are almost always read in an upright orientation and frequently occur in closely spaced strings (i.e., words). Further, precisely recognizing both the identity and relative location of letters in words is imperative during reading (e.g., the strings “salt” and “slat” have completely different meanings). This massive but highly constrained reading experience may cause changes in how the visual system processes letters, such that crowding is minimized for upright characters. Second, reading relies almost entirely on perceiving crowded information in the fovea and parafovea, whereas many other types of natural vision do not require individuating a large number of discrete elements that are so closely spaced near the fovea. As a result, crowding for upright letters may be particularly reduced near the point of fixation. Third, readers of scripts with a left-to-right organization, such as English and Spanish, rely heavily on parafoveal information from the right side of fixation to control their eye movements (Schotter, Angele, and Rayner, 2012), to guide recognition of currently fixated words (Veldre & Andrews, 2016), and in some cases to recognize upcoming words without fixating them (Rayner et el., 2011). Consequently, readers of such languages may develop mechanisms that specifically minimize crowding to the right of fixation.

There is already some evidence that crowding effects differ between letters and other types of stimuli. For example, recognition accuracy is better for crowded letters than for crowded symbols (Grainger, Tydgat, and Issele, 2010), and this advantage has been shown to be true particularly for horizontally oriented arrays rather than vertically oriented arrays (Vejnović and Zdravković, 2015). However, these experiments compared letters with symbols or digits (e.g. Tydgat and Grainger, 2009; Grainger, Tydgat, and Issele, 2010; Vejnović and Zdravković, 2015; Winsler, Grainger, and Holcomb 2022), making it possible that the differences in accuracy may be due to differences in the physical stimuli themselves rather than differences in experience with the stimuli. Additionally, almost all studies used behavioral accuracy at a single target-flanker spacing as the dependent measure, potentially reflecting post-perceptual processes rather than crowding itself (but see Experiment 5 in Grainger, Tydgat and Issele, 2010). Moreover, no prior research has examined whether individual differences in reading experience are associated with individual differences in crowding. Thus, prior research does not clearly address the question of whether massive visual experience with reading impacts the mechanisms that underlie crowding.

The present studies

The present studies directly addressed this issue using two preregistered, converging approaches. In Chapter 1, we present an experimental approach in which we compared crowding for upright letters with crowding for less familiar stimuli (referred to as Study 1a and 1b). In

Chapter 2, we present an individual-differences approach in which we asked whether specific attributes of crowding are related to proxy measures of lifetime reading experience (referred to as Study 2). As illustrated in Figure 2, both experiments used multiple target-flanker distances to establish the target-flanker distance at which crowding occurred. We refer to this as the *crowding threshold* (also known as critical distance), defined as the target-flanker distance that yielded a target discrimination accuracy of 82% correct. This allowed us to determine whether familiarity impacts crowding per se rather than impacting overall accuracy. We also tested multiple target eccentricities on both the left and right sides of fixation so that we could determine whether the spatial profile of crowding is different for upright letters (i.e., decreased crowding for upright letters near fixation and in the right visual field). Note that we focused here on radial crowding along the horizontal meridian, which is the primary type of crowding experienced during reading of left-to-right scripts.

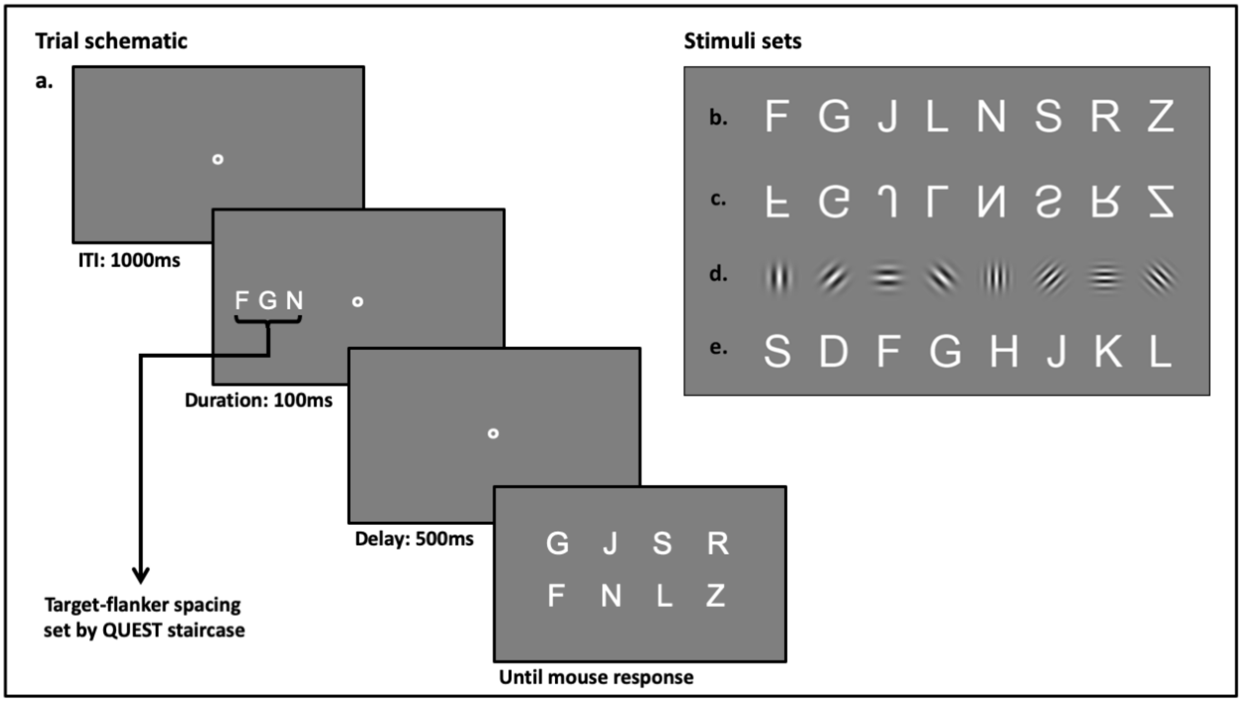

Figure 2. **a.** Example of a trial with upright letters in Study 1. The target was the central item in each three-item string of letters. After a period of fixation on the central fixation point, a target and two flankers appeared for 100 ms. The target was unpredictably placed at one of three eccentricities in the left or right visual field. After a 500 ms of delay, an array of eight choice letters appeared and the participant used the mouse to click on the letter that matched the target. **b.** Upright letter stimuli used in Studies 1a and 1b. **c.** Inverted letter stimuli used in Study 1a. **d.** Gabor patch stimuli used in Study 1b. **e.** Upright letter stimuli used in Study 2.

In Study 1a, we assessed the effect of experience by comparing crowding for upright letters with crowding for inverted letters. Inverted letters are much less familiar than upright letters, but they provide a much better control for low-level stimulus properties than other stimuli such as symbols. For example, inverting a letter does not change perimetric complexity (Pelli et al., 2006), spatial frequency content, or basic visual features. This makes it possible to be nearly certain that any differences in crowding between upright and inverted letters were caused by differences in experience rather than by intrinsic differences in visual properties. We tested three predictions: a) that the visual system can tolerate closer target-flanker distances for upright letters than for inverted letters; b) that the advantage for upright letters would be particularly strong near fixation; and c) that this advantage would be greater to the right of fixation than to the left. Study 1b tested the same predictions by comparing upright letters with Gabor patches, which are unfamiliar but closely match the tuning properties of neurons in primary visual cortex (Jones & Palmer, 1987). The primary goal of including the Gabor patch condition was to measure crowding over spatial locations with a commonly used simple stimulus to compare with letters. We predicted that the differences in crowding between upright letters and Gabor patches would be even stronger than the differences between upright and inverted letters. This could be due to a greater difference in experience or to a greater difference in low-level stimulus properties.

Method

For Study 1, the procedure, sampling plan, and analysis plan were preregistered at https://osf.io/3fpuw/?view_only=aed387d6dff247d5ac0ee6694dd999f4. Stimulus presentation scripts, data, and analysis scripts for Study 1 are available at https://osf.io/q32xg/?view_only=43499a011a75405c837d4a5d8865715c.

Participants

The participants were undergraduate students enrolled in psychology courses at University of California, Davis. All participants used for analysis were native and current readers of a Latin script (typically English or Spanish), native being defined as learned as their first written language or concurrently with their first written language. All participants reported having normal or corrected-to-normal vision. On the basis of a power analysis of data from a pilot experiment similar to Study 1a (N = 57,) we preregistered a sample size of 60 usable participants for Studies 1a and 1b (see preregistration). Participants were considered usable if they met minimal standards for on-task performance as described below. All studies were approved by the Institutional Review Board at University of California, Davis, and all participants provided informed consent. In Study 1a, the average age of the 60 participants was 19.57 years (SD = 1.65). In Study 1b, the average age was 19.24 years (SD = 2.1). Additional demographic information for participants can be found in Table 1, which uses standard US-based categories for race and ethnicity (US Office of Management and Budget, 2016).

Demographics of Participants

Characteristic	Study 1a (n =60)		Study 1b (n = 60)		Study 2 (n = 250)	
	n	*%*	*n*	*%*	*n*	*%*
Gender						
Female	47	78.3	45	75	56	22.4
Male	11	18.3	12	20	186	74.4
Other gender or not reported	2	3.3	3	5	8	3.2
Race						
Asian	27	45	27	45	119	47.6
White	23	38.3	20	33.3	78	31.2
Mixed race	4	6.6	5	8.3	15	6
Black	2	3.3	0	0	8	3.2
Native American	2	3.3	0	0	5	2
Other race or not reported	2	3.3	8	13.3	25	10
Ethnicity						
Not Hispanic or Latino	42	70	41	68.3	174	69.6
Hispanic or Latino	18	30	15	25	74	29.6
Not reported	0	0	4	6.6	2	0.8

Table 1. Gender, Race, and Ethnicity data for participants in Study 1a, Study 1b, Study 2.

Materials and procedure

Stimulus presentation and data collection were controlled using PsychoPy (Pierce, 2007). Stimuli were presented on a HP ZR2440w LCD display with a gray background (26.8 cd/m^2) at a nominal viewing distance of 100 cm.

The stimuli and procedure are illustrated in Figure 2. Studies 1a and 1b were identical except that Study 1a compared upright letters (highly familiar stimuli) with inverted letters (less familiar stimuli), whereas Study 1b compared the same upright letters (highly familiar stimuli) with Gabor patches (unfamiliar stimuli). In each study these two stimulus classes were factorially combined with 6 target locations on the visual field: 2, 4, and 6 degrees of visual angle to the left and right of fixation (on the horizontal meridian). The familiar upright letter stimuli were selected from a set eight white letters (F, G, J, L, N, S, R, Z) drawn in Arial font with a height of 1.0° and a luminance of 109 cd/m^2 (Figure 2b). In Study 1a, the less familiar inverted letter stimuli were these same letters inverted (i.e., flipped vertically, not rotated 180 degrees – Figure 2c). These specific letters were chosen to ensure that their upright and inverted forms would be different from each other. In Study 1b, the unfamiliar Gabor patch stimuli were selected from a set of eight Gabor patches (Figure 2d) with a spatial standard deviation of ± 0.33° that varied in orientation (0°, 45°, 90°, or 135° from upright) and spatial frequency (4 or 6 cycles/°). In pilot testing, we found that these eight combinations of orientation and spatial frequency led to performance levels that were comparable to the eight inverted letters (which was not the case if we used eight different orientations with a single spatial frequency). Note that crowding is largely independent of the size of the target and flankers (Tripathy & Cavanagh, 2002), so we held stimulus size constant across eccentricities.

The task for each experiment was to identify the central stimulus in the array of three stimuli presented in each trial. For each of the twelve combinations of location and stimulus type, 50 trials were run, shuffled such that each condition was approximately evenly distributed across the 600 total trials. Each trial began with the subject fixating on a circular fixation point in the center of the screen, followed by a target appearing for 100 ms. The target was flanked horizontally by nontarget stimuli from the same set (e.g., nontarget inverted letters if the target was an inverted letter). The target-flanker distance was determined by a staircase procedure, described below. After 500 ms, a response array appeared, with the 8 possible stimuli from the stimulus set for that trial (e.g., the 8 inverted letters if it was an inverted trial), randomly distributed in a 4x2 grid on the screen. Subjects used a mouse to click on the stimulus that matched the target from that trial. After response, the fixation stimulus reappeared, and a new trial started after 1 second. For each experiment, participants completed a 24-trial practice version of the task before completing the actual experiment.

In each of the 12 conditions, the center-to-center distance between the target and each flanker was set according to the recommended "intensity" of a QUEST staircase (Watson & Pelli, 1983), targeting a recognition accuracy of 82% correct for that condition. For all locations and stimulus conditions, the staircase had a prior of 0.5φ (i.e. spacing of 50% of the target's eccentricity, φ) with an SD of .3φ. This prior was centered at 0.5φ because this represents the traditional estimate of the crowding distance (Bouma, 1970). Note that this measure of spacing abstracts away from the raw target-flanker distances and incorporates the linear eccentricity scaling that is fundamental to crowding ("Bouma's law"; Pelli and Tillman, 2008). For example, both a raw crowding threshold of 1° at an eccentricity of 4° and a raw crowding threshold of 0.5° at an eccentricity of 2° would each correspond to a scaled crowding threshold of 0.25φ. Thus, the null hypothesis for our eccentricity manipulation was that the scaled crowding threshold would be constant across eccentricities. The range of the QUEST staircase was limited to sample values between a maximum spacing of 1 and a minimum spacing of 0. The beta (shape) parameter for the staircase was fixed at 3.5, the gamma (lower asymptote) parameter was fixed at 0.125 (i.e., 1 divided by 8 possible answer choices), and the delta (upper asymptote) parameter was fixed at 0.02.

As a quality control measure, each session also contained 48 catch trials in which a target was presented without any flankers. These trials were distributed unpredictably across the session and included three trials for each of the 16 possible target stimuli (i.e. 3 for each letter and inverted letter/Gabor patch), with eight trials at each of the six locations. The catch trials were designed to be easy, so below-ceiling accuracy on these trials likely indicates either insufficient visual acuity or frequent lapses of attention. Participants were considered unusable if they missed more than two of the 48 catch trials. Participants were also considered unusable if more than two of the 12 staircases ended up at the maximum flanker distance (a crowding threshold of greater than 95% spacing, which is implausible for individuals with typical vision). These exclusion criteria resulted in 7 unusable subjects in Study 1a, and 36 unusable subjects in Study 1b. Unusable participants were simply replaced so that we could reach the target N of 60 usable participants in each study. Individual staircases which were estimated at near the maximum value were also excluded from analysis (0.83% of staircases in Study 1a and 3.5% of staircases in Study 1b). All data exclusion procedures were specified in the preregistration.

Data analysis

The dependent variable was the *crowding threshold*, defined as the target-flanker spacing required to achieve 82% accuracy (expressed as proportion of the target's eccentricity). This value was computed as the mean of the posterior distribution for the 50-trial QUEST staircase for each condition. Hence, each participant contributed 12 values to the analysis – one crowding threshold for each combination of the two stimulus conditions (familiar and unfamiliar), three eccentricities (2°, 4°, and 6°), and two hemifields (left and right).

These data were modeled separately for each experiment using linear mixed effect regression as implemented in the lme4 R package (Bates, et al., 2014). The structure of the model included fixed effects for Eccentricity, Hemifield, and Familiarity, and all possible interactions between these three variables. Hemifield and Familiarity were effect coded (i.e., -0.5 and 0.5 for left vs. right; -0.5 and 0.5 for familiar vs. unfamiliar). Eccentricity was coded as a continuous variable centered at the intermediate eccentricity (i.e., 2°, 4°, and 6° were coded as -0.5, 0, and 0.5, respectively). The random effect structure included random intercepts for Subject, as well as by-subject random slopes for the effect of each of the fixed effects. In other words, the model did not assume that the slopes were identical across participants. As specified in the preregistration, the random slopes for the interactions between the fixed effects were dropped from all models to allow them to converge. Significance for each effect was tested with a Type 3 Wald chi-square test using Satterthwaite's method through the lmerTest package (Kuznetsova, Brockhoff, and Christensen, 2017). These tests were preregistered and used uncorrected *p* values with an alpha of .05. Beta weights and standard errors are also reported from the models, their units being spacing as a ratio of the target's eccentricity (φ)

To aid the interpretation of interaction effects, follow-up analyses were conducted by fitting the same model separately for each stimulus type without the fixed or random effects of Familiarity. Additionally, a follow up model was fit to compare the inverted letters from Study 1a with the Gabor patches in Study 1b. This model had the same structure as the main models, but without random slopes for Familiarity (because Familiarity was a between-subjects variable for this analysis).

When an error is made during a crowding task, participants often report the identity of a flanker. This is known as a substitution error or source confusion. The present study was not designed to analyze substitution errors, but for the curious reader, an exploratory analysis of the substitution errors from Study 1 is available in the supplementary materials.

Results and Discussion

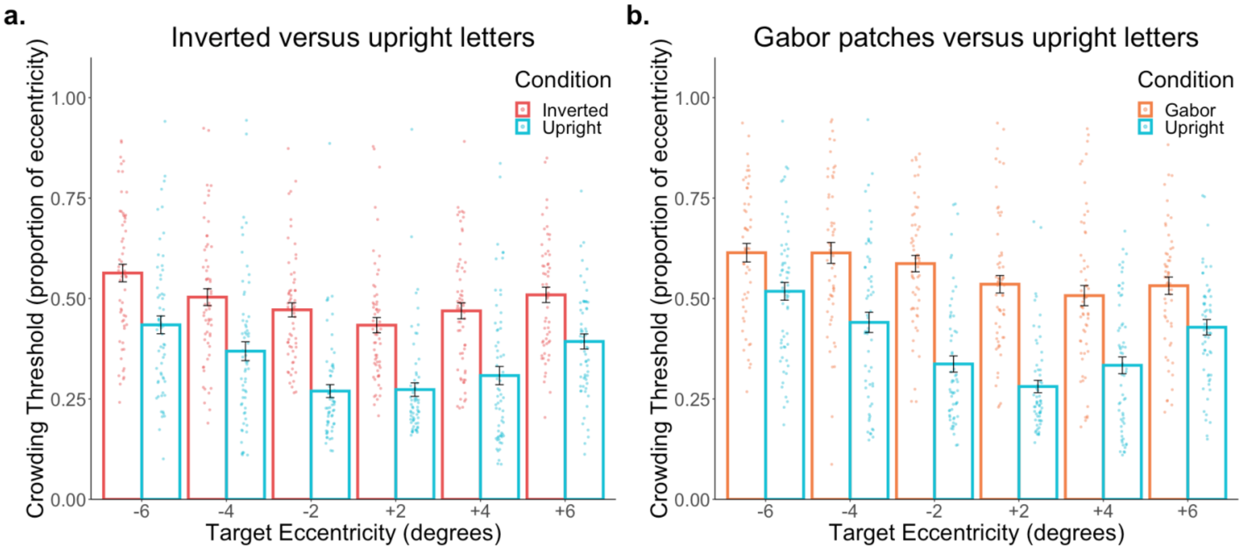

Figure 3. Mean eccentricity-scaled thresholds for each location and familiarity condition, comparing upright and inverted letters in Study 1a (**a**) or upright letters and Gabor patches in Study 1b (**b**). Error bars show ±1 SEM. Dots show single-participant thresholds. Note that a perfectly linear relationship between eccentricity and crowding threshold would lead to a flat function over eccentricities (e.g., all values ~ 0.5φ).

Study 1a: Upright versus inverted letter crowding

Figure 3a shows the mean eccentricity-scaled crowding thresholds for each cell of Study 1a. The crowding threshold was lower for upright letters than for inverted letters across the locations tested ($\beta = 0.155$, SE = 0.009, $p < 0.001$). Thus, although the upright and inverted letters were essentially identical in terms of their visual properties, the upright letters could be identified with closer spacing than the inverted letters. This provides evidence for an effect of experience on the most fundamental property of crowding, namely its dependence on target-flanker distance.

Crowding thresholds were lower when the target was closer to fixation than when the target was farther away ($\beta = 0.112$, $SE = 0.011$, $p < 0.001$), despite the fact that the thresholds were already scaled to be proportional to the eccentricity. The eccentricity effect was steeper for the upright letters than for the inverted letters, leading to a significant Familiarity x Eccentricity interaction ($\beta = -0.055$, $SE = 0.016$, $p < 0.001$). Thus, the second most fundamental property of crowding—its eccentricity function—also differed between highly familiar upright letters and less familiar inverted letters.

We decomposed the Familiarity x Eccentricity interaction by testing the eccentricity effect separately for upright letters and for inverted letters. The Eccentricity effect was significant for both upright letters ($\beta = 0.139$, $SE = 0.014$, $p < 0.001$) and inverted letters ($\beta = 0.083$, $SE = 0.011$, $p < 0.001$) when they were analyzed separately. Thus, even though the eccentricity effect was greater for upright than inverted letters, the eccentricity-scaled thresholds increased with eccentricity for both stimulus types. This is a violation of Bouma's law, which specifies that the eccentricity-scaled threshold should be constant (i.e., that the absolute threshold should be a constant proportion of the eccentricity).

As illustrated in Figure 3a, crowding thresholds were significantly smaller for targets to the right of fixation than to the left ($\beta = -0.036$, $SE = 0.012$, $p = 0.005$). This was predicted because experienced readers of Latin scripts regularly sample letter information to the right of fixation prior to each saccade. However, the Familiarity x Hemifield interaction was near zero and not statistically significant ($\beta = -0.008$, SE = 0.013, $p = 0.564$), suggesting a general advantage for the right visual field rather than a specific rightward bias for upright letters.

Study 1b: Upright letter versus Gabor patch crowding

Figure 3b shows the mean eccentricity-scaled crowding thresholds for each cell of Study 1b. Upright letters had smaller crowding thresholds than Gabor patches across the locations tested ($\beta = 0.183$, SE = 0.016, $p < 0.001$). This provides additional evidence that the visual system can tolerate closer spacing for more familiar stimuli (upright letters) than for less familiar stimuli (Gabor patches).

The Gabor patches used in Study 1b are very rare in the natural environment, whereas the inverted letters used in Study 1a are presumably at an intermediate level of familiarity between Gabor patches and upright letters. We therefore compared the data from the inverted letters from Study 1a with the data from the Gabor patches from Study 1b. We found that thresholds were significantly lower for the inverted letters than for the Gabor patches ($\beta = 0.077$, SE = 0.022, $p < 0.001$). These results suggest a graded effect of experience, with the lowest thresholds for the most familiar stimuli (upright letters), intermediate thresholds for moderately familiar stimuli (inverted letters), and the highest thresholds for highly unfamiliar stimuli (Gabor patches).

As in Study 1a, eccentricity-scaled thresholds for upright letters in Study 1b decreased substantially as target eccentricity decreased, violating Bouma's law. However, Gabor patches followed Bouma's law, exhibiting little or no effect of eccentricity on the eccentricity-scaled thresholds. This pattern led to a significant Familiarity x Eccentricity interaction ($\beta = -0.149$, SE = 0.023, $p < 0.001$). When analyzed separately, the effect of Eccentricity was significant for letters ($\beta = 0.165$, SE = 0.016, $p < 0.001$) but not for Gabor patches ($\beta = 0.016$, SE = 0.018, $p = 0.386$).

Additionally, an across-experiment follow-up analysis indicated that the effect of eccentricity was significantly greater for inverted letters than for Gabor patches (Familiarity × Eccentricity interaction: $\beta = -0.067$, SE = 0.022, $p = 0.002$). In other words, Bouma's Law was confirmed for Gabor patches, but it was violated for inverted letters and violated even more strongly for upright letters. These results mirror the pattern of effects observed for the overall thresholds, with the largest eccentricity effect for the highly familiar letters, weaker eccentricity scaling for the less familiar inverted letters, and no eccentricity scaling beyond Bouma's law for the highly unfamiliar Gabor patches.

Finally, crowding thresholds were lower for targets in the right visual field than for targets on the left ($\beta = -0.088$, SE = 0.016, $p < 0.001$). However, the interaction between Familiarity and Hemifield was near zero and not significant ($\beta = -0.007$, SE = 0.019, $p = 0.691$), consistent with a general advantage for the right visual field.

Chapter 2: Individual differences in visual crowding

In Study 1, we found several effects consistent with the hypothesis that reading experience leads to changes in fundamental aspects of visual crowding. Crowding thresholds were lower for upright letters than for inverted letters, and lower for both upright and inverted letters than for Gabor patches. This parallels the amount of familiarity we would expect for these three stimulus classes. In addition, whereas Gabor patches followed Bouma's law, with no change in eccentricity-scaled thresholds across eccentricities, these thresholds were reduced near fixation for letters (especially upright letters). This fits with the fact that reading is primarily performed at or near fixation. Finally, crowding was reduced in the right visual field (for all stimulus types), matching the fact that information from the right visual field plays an important role in guiding saccades during the reading of languages that use the Latin script. Although some of these effects might be independent of reading experience, the differences between upright and inverted letters in overall thresholds and eccentricity effects were almost certainly a result of reading experience given that inverted letters are almost identical to upright letters in every property other than experience.

In Study 2, we aimed to provide additional evidence that crowding for upright letters is related to reading experience, using a large-N individual-differences approach. Specifically, we asked whether inter-subject variability in reading experience (as assessed by two proxy measures) is associated with the putative reading-related properties of crowding that we observed for upright letters in Study 1 (reduced thresholds, a drop in eccentricity-scaled thresholds near fixation, and a right hemifield advantage). Our primary proxy measure—the Andrews spelling test—quantifies the ability to recognize misspelled words and is thought to reflect the quality or precision of word representations, which are developed through reading practice (Andrews, Veldre, and Clarke, 2020). Our secondary proxy—the author recognition task—quantifies the number of author names that an individual recognizes, which is an indirect measure of lifetime reading experience (Acheson, Wells, and MacDonald, 2008). We predicted that individuals who scored higher on these tests would have decreased crowding thresholds, especially for locations closer to fixation and locations in the right visual field.

Additionally, we measured visual working memory capacity with a change localization task that provides a sensitive and reliable measure of general visual and cognitive ability (Johnson et al., 2013). This allowed us to assess the possibility that the relationship between our reading experience measures and crowding reflects some broader aspect of visual cognition rather than being specific to reading.

Method

For Study 2, the procedure, sampling plan, and analysis plan were preregistered at https://osf.io/fpwv3/?view_only=628666eff3744ff6994bb154c87d1a7c. Stimulus presentation scripts, data, and analysis scripts for Study 2 are available at https://osf.io/qabh6/?view_only=c4c8577c06e9404a913f4b964bc0d0b5.

Participants

The original preregistered sampling plan had the same inclusion criteria as in Study 1, including all individuals who were native and current readers of any language using a Latin script. Based on a power analysis using data from 115 participants in pilot study, we preregistered a sample size of at least usable 200 participants, with data collection continuing until the end of the spring 2022 term (see preregistration). This led to a sample of 267 usable subjects. However, we later realized that our measures of reading experience—which are based on English-language spelling ability and the recognition of English-language authors—were designed for native readers of English and have not been validated for native readers of other languages. We therefore limited inclusion to native and current readers of English (N = 250) and excluded 17 individuals who were native and current readers of other languages that use a Latin script (primarily Spanish). This difference did not impact the statistical significance of any of the effects, but it is nonetheless the more appropriate sampling method, so the analyses provided here were limited to the 250 native readers of English. Analyses of the full sample of 267 participants are provided in Appendix B (see Tables S4 and S5). The average age of the 250 participants was 20.04 years (SD = 2.27). Additional demographic information can be found in Table 1.

Materials and procedure

All participants performed four tasks in a fixed order: a spelling task, an author recognition task, a crowding task, and a visual working memory task. The spelling and author recognition tasks were run first to ensure that they would not be contaminated by an individual's perceived performance on the crowding task. The working memory task was run last because it was considered secondary. In addition, it was important to keep the order constant across participants to avoid adding an uncontrolled source of variance among individuals.

Crowding task

The crowding task for Study 2 was identical to that for Study 1 except as follows. Only familiar stimuli (upright letters) were used. The set of letters used was the eight consonants on the middle row of a standard keyboard (S, D, F, G, H, J, K, L). Rather than clicking with a mouse to report the perceived target letter, participants responded by pressing the corresponding key on the keyboard. To increase the precision of the single-participant crowding thresholds, the number of trials for each condition was increased from 50 to 70. Consequently, the total number of trials for the crowding task in Study 2 was 420. Participants completed an 18-trial practice version of the task before completing the actual experiment. A total of 42 catch trials were presented, including seven trials at each of the six locations. As specified in the preregistration, participants were considered unusable if they missed more than two of these 42 catch trials or if more than one of the six staircases ended up at the maximal flanker distance. These exclusion criteria resulted in 39 subjects being categorized as unusable and therefore being replaced. Individual staircases which were excluded for being near the maximum value amounted to 0.6% of staircases in Study 2.

Visual working memory task

Visual working memory capacity was measured using a color change localization task with a set size of 6 items, presented on the same video monitor used for the crowding task. Change localization performance is substantially more reliable than performance of the more common change detection task (Kyllingsbaek & Dundesen, 2009; Zhao, Vogel, and Awh, 2022), is strongly correlated with change detection performance (Zhao, Vogel, and Awh, 2022), and is strongly correlated with IQ and other measures of overall cognitive ability (Johnson et al., 2013). For each trial in this task, six colored disks (each with a radius of half a degree of visual angle) appeared for 200 ms, equally spaced around a notional circle with an eccentricity of 3° from a central fixation cross. After a retention period of 900 ms, the disks reappeared, with a disk at one randomly selected location having changed color. The participants then used a computer mouse to click on the changed disk. The colors of the disks were randomly drawn from sets of approximately isoluminant colors, such that the colors in a given array were roughly equally spaced from across the color wheel. The color of the changed disk was 120 degrees on the color wheel away from the original color. Participants completed 96 trials after experiencing 12 practice trials. The dependent measure for this task was the number of trials with a correct response, which is perfectly correlated with the traditional K measure of storage capacity for this task (Johnson et al., 2013). The split-half reliability of this task in our sample of participants was high ($r = .80$).

Spelling task

Our primary proxy for reading experience was the Andrews spelling task, which measures a participant's ability to identify misspelled words and has been shown to have high test-retest reliability ($r = .93$; Andrews & Hersch, 2010). Spelling ability is thought to represent the quality or precision of a person's lexical representations which are learned through experience (for a review see Andrews, Veldre, and Clarke, 2020). We used a recently updated version (Andrews, Veldre, and Clarke, 2020). Participants were given a sheet of paper with 88 words arranged in a grid. Half of the words were misspelled, and the participant was instructed to circle these words. The score was the number of circled misspelled words plus the number of uncircled correctly spelled words.

Author recognition task

Our secondary proxy for reading experience was the *author recognition task*. This task is designed to quickly measure print exposure, based on the assumption that people who have read more will be able to recognize more authors (Stanovich and West 1989). We used an updated version (Acheson, Wells, and MacDonald, 2008), in which 130 real and fake author names were intermixed in a grid on a sheet of paper. Participants circled all the names they recognized as being authors. Participants received one point for each correct author circled and were penalized one point for each non-author circled. Meta-analysis of studies using this task show it has good reliability (Cronbach's α = .75-.89; see Mol and Bus, 2011). However, note that this task is not a very good measure for people who come from non-Western backgrounds (McCarron and Kuperman, 2021), like many of the students at University of California, Davis. Consequently, we did not expect it to be as good of an indicator of reading experience as the Andrews spelling task.

Data analysis

The analysis strategy was patterned after that used in Study 1. Crowding thresholds at each of the 6 locations were modeled with fixed effects for Hemifield, Eccentricity, and the interaction between the two. Additionally, and interacting with each of these effects, there were fixed effects for the three individual difference measures –visual working memory, spelling, and author recognition. Hence the highest order interactions were 3-way interactions between Hemifield, Eccentricity, and each of the 3 individual difference measures (as specified in the preregistration). Each of the individual difference measures was normalized across the sample of subjects, putting all three measures onto the same scale of z scores. In the present sample, the correlation between spelling and author recognition was 0.328, the correlation between visual working memory and spelling was 0.109, and the correlation between visual working memory and author recognition was -0.016. The random effect structure included random intercepts for subjects and by-subject random slopes for Hemifield and Eccentricity. Significance testing was identical to the models in Study 1. The full model output is available in Appendix B (see Table S4).

As with Study 1, an analysis of the substitution errors made in Study 2 and how their likelihood is predicted by scores on the individual difference measures is available in Appendix A.

Results and Discussion

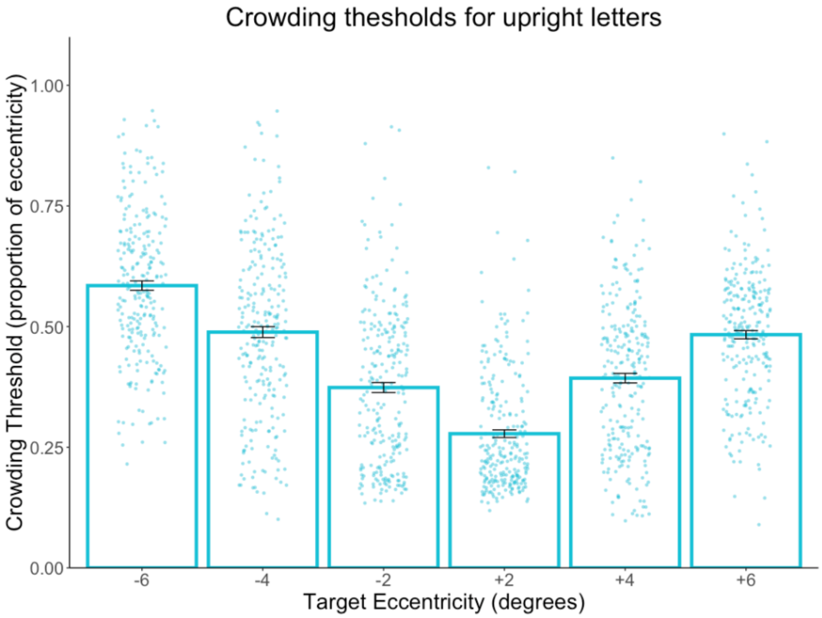

Figure 4. Mean eccentricity-scaled thresholds for upright letters at each location in Study 2. Error bars show ±1 SEM. Dots show single-participant thresholds. Note that a perfectly linear relationship between eccentricity and crowding threshold would lead to a flat function over eccentricities (e.g. all values ~ 0.5φ).

As shown in Figure 4, the eccentricity-scaled crowding thresholds were significantly smaller when the targets were closer to fixation ($\beta = 0.211$, SE = 0.007, $p < 0.001$) and were significantly smaller for targets to the right of fixation than for targets to the left ($\beta = -0.099$, SE = 0.008, $p < 0.001$). This replicates the pattern observed for upright letters in Study 1.

To visualize the correlations between crowding thresholds and the individual difference measures, we plotted the estimated slope of the relationship between eccentricity-scaled crowding thresholds and the spelling task (Figure 5a), the author recognition task (Figure 5b), and the visual working memory task (Figure 5c), averaged over the 6 locations. A negative slope indicates less crowding (a lower threshold) for people with higher scores on these variables. To visualize the effect of each individual difference measure over the 6 locations, we plotted the model-predicted change in crowding threshold for an increase of one standard deviation on the spelling task (Figure 5d), the author recognition task (Figure 5e), and the visual working memory task (Figure 5f), for each of the 6 locations. In Figure 5 g-i, we also plotted these same model-predicted changes for models fit on each individual difference measure separately, which show the effects of each variable without controlling for the effects of the other individual difference variables. The results of these models had the same pattern of significance as the main model and are provided in the supplementary materials (see Tables S3–S5 in Appendix A).

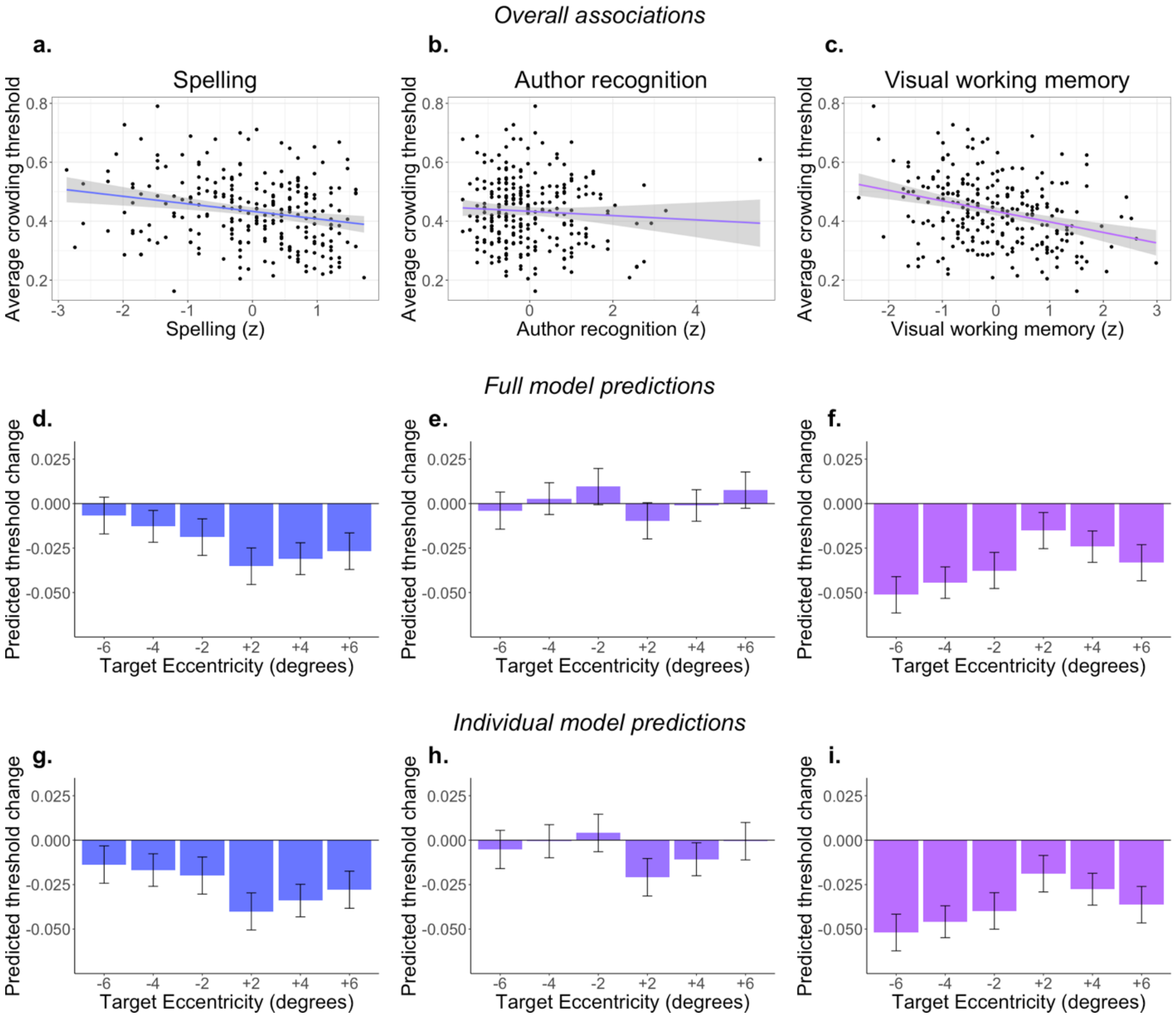

Figure 5. a-c. Relationship between average crowding threshold (over the 6 locations) and each of the individual difference measures (Spelling, Author recognition, and Visual working memory). Dots are individual subjects, and the lines are simple linear regression fits. The shaded region shows the standard error of estimate for the regression line at each point. **d-f.** Model-predicted change in crowding threshold for an increase of 1 standard deviation on each of the individual difference measures, per location, in the full model. **g-i.** Model-predicted change in crowding threshold for an increase of 1 standard deviation on models fit with each of the individual difference measures separately. Error bars show ±1 SEM.

Figure 5a shows that higher scores on the spelling test were associated with significantly lower crowding thresholds across the six locations ($\beta = -0.022$, SE = 0.007, $p = 0.003$). Figure 5d shows that this effect was especially prominent when the target was presented to the right of fixation, leading to a Spelling × Hemifield interaction ($\beta = -0.018$, SE = 0.008, p = 0.027). These results provide converging evidence for the hypothesis that greater reading experience and lexical precision is associated with reduced crowding, allowing more closely spaced letters to be identified, especially in the right hemifield where English readers have the most experience recognizing crowded letters. Note that performance on the spelling test had a numerically larger impact for stimuli closer to fixation, but the interaction with Eccentricity was not significant ($\beta = 0.010$, SE = 0.007, p = 0.164). Thus, we found that greater spelling ability—which is a proxy for reading experience—significantly predicted reduced overall crowding thresholds and a greater right hemifield advantage. It is possible the association between Spelling and Eccentricity would have reached significance if we had included a broader range of eccentricities or if we had an even larger sample of subjects.

As shown in Figure 5b, performance on the author recognition task did not predict overall crowding thresholds ($\beta = 0.001$, SE = 0.007, p = 0.904). However, there was a significant three-way interaction of Author Recognition × Hemifield × Eccentricity ($\beta = 0.031$, SE = 0.012, p = 0.013). As shown in Figure 5e, this interaction appeared to reflect increased scores on author recognition being associated with lower crowding thresholds for targets specifically in the right hemifield and at the nearest eccentricity. One reason for the limited effect of author recognition may be that this measure was positively correlated with spelling ability ($r = .35$ in our sample), and variance in crowding was attributed to spelling ability. Consistent with this, when spelling was removed from the model, the effect of author recognition increased markedly (but was still focused on the locations immediately to the right of fixation – see Figure 5h).

To examine whether crowding is related to reading experience per se or whether associations would be seen for any measure of cognitive ability, we assessed whether individual differences in visual working memory performance showed the same pattern of association with crowding thresholds. As illustrated in Figure 5c, greater visual working memory scores were associated with significantly lower overall crowding thresholds ($\beta = -0.034$, SE = 0.007, $p < 0.001$). However, as shown in Figure 5f, the spatial pattern was the opposite of the effects of spelling ability. Specifically, the visual working memory effects were larger to the left of fixation ($\beta = 0.020$, SE = 0.008, $p = 0.010$) and larger at greater eccentricities ($\beta = -0.016$, SE = 0.007, $p = 0.024$). Thus, although both visual working memory performance and spelling performance were associated with decreased crowding thresholds, each explained unique variance and had a different spatial pattern, suggesting different underlying mechanisms.

Discussion of Studies 1 and 2

Letters are tightly packed within words, and the reduction in perceptibility produced by this crowding is a major limitation on reading (Pelli et al., 2007). The results of Studies 1 and 2 demonstrate that the human visual system minimizes the deleterious effects of crowding in three ways that match the specific demands of reading. First, the visual system can tolerate closer spacing for upright letters than for inverted letters or Gabor patches, which matches the fact that almost all reading is done with upright letters. Second, the effects of crowding on upright letters are increasingly minimized for stimuli closer to fixation, which matches the fact that reading puts high demands on the identification of tightly spaced but highly similar stimuli near the fovea. Third, the visual system can tolerate closer spacing of stimuli in the right visual field than in the left visual field, which matches the fact that readers of Latin scripts make extensive use of parafoveal information in the right visual field (Schotter, Angele, and Rayner, 2012; Veldre & Andrews, 2016; Rayner et el., 2011). In addition, people who scored higher on the Andrews spelling test—which appears to reflect experience-driven changes in the precision of visual word representations (Andrews, Veldre, and Clarke, 2020)—could tolerate closer spacing for upright letters, especially in the right visual field.

The first two of these effects are almost certainly caused by the massive experience with reading that most adults in industrialized societies have obtained. Although it is not practical to experimentally manipulate experience at this scale in humans, the observed differences in crowding between upright and inverted letters would be difficult to explain in any other way. That is, given that upright and inverted letters are almost identical in terms of their physical properties, it is implausible that people without experience reading the Latin script would show different crowding patterns for upright and inverted letters from this script. Thus, the results of Study 1 provide strong evidence that reading experience produces changes in fundamental aspects of visual processing. An open question is whether similar changes in visual processing are produced by experience with scripts that are very different from the Latin script, such as the much more complex logograms used in Chinese languages, or in languages read from right-to-left such as Arabic.

Across Studies 1a and 1b, we also found evidence for a graded effect of experience on both crowding thresholds and eccentricity effects, with moderately familiar stimuli (inverted letters) producing effects that were intermediate between those for highly familiar stimuli (upright letters) and highly unfamiliar stimuli (Gabor patches). Gabor patches obeyed Bouma's Law (Pelli and Tillman, 2008), which states that the critical distance for crowding remains constant when expressed as a proportion of target eccentricity. Inverted letters deviated significantly from Bouma’s Law—producing lower-than-expected thresholds near fixation—and upright exhibited this deviation even more strongly. This may indicate that at least some of the mechanisms responsible for crowding are fundamentally altered for letters, or that specialized attentional mechanisms are available for letters but not for inverted letters or Gabor patches. Further research would be needed to determine whether the differences between letters and Gabor patches are a result of experience or a result of the physical properties of the stimuli. For instance, crowding distances for Gabor patches may be greater because of increased target-distractor similarity. This may also be responsible for the more-typical pattern of eccentricity scaling observed for Gabor patch crowding, due to information pooling within receptive fields of early visual cortex, which is indeed the primary mechanism in many dominant models of visual crowding (e.g. Levi, 2008, Pelli and Tillman, 2008). However, the low-level similarity structure was highly controlled between the upright and inverted letters used in Study 1a. Hence, the difference in both crowding distance and eccentricity scaling we observed between upright and inverted letters is hard to explain in models of crowding which rely solely on irreversible information loss due to pooling in early visual processing. Our results fit better with models of crowding that allow for crowding at multiple levels in the visual hierarchy (e.g., Manassi and Whitney, 2018). One such model is the hierarchical sparse selection model (Chaney, Fischer and Whitney, 2014) in which crowding for objects is a function of the receptive fields of the populations of neurons that represent the object itself. This model predicts differences in crowding across stimulus types, and accounts for other related results such as the finding that upright faces are crowded more by other upright faces than inverted ones (Louie, Bressler and Whitney, 2007). Under this framework, the constraints of reading may spatially bias the tuning of letter representations, such that their receptive fields are smaller, particularly closer to the fovea. The finding that inverted letters also deviate from Bouma's law may indicate that some of this tuning may be occurring at the level of letter features as well. In any case, a complete model of visual crowding must incorporate some mechanism to allow altered patterns of crowding for highly familiar stimulus classes.

Although conclusions about causation are more difficult for the correlational approach used in Study 2, the observed association between spelling performance and crowding thresholds, particularly in the right visual field, is at least consistent with the proposal that reading experience alters letter crowding. Performance on the author recognition test was not significantly related to overall crowding thresholds, although it was significantly associated with lower crowding thresholds in the near-right parafovea, a key location for readers of Latin scripts. Hence, this effect is also consistent with a role of reading experience on crowding. These effects were not driven by greater overall cognitive ability in people with high scores on the spelling test, because a very different pattern of associations was observed for an independent measure of working memory capacity. While we suggest there is good reason to believe that reading experience would cause changes in letter crowding, there is also the possibility that the causal direction is flipped or mediated by a third variable. That is, having lower crowding thresholds for letters (perhaps due to some factor unrelated to reading) may lead people to spend more time reading, leading to the correlations we observed.

We would like to stress that although our proxies for reading experience are widely used, they are far from perfect measures of actual reading experience. Reading is a complex human activity that involves the integration of many visual and cognitive skills. Spelling ability is thought to index the quality or precision of visual word representations which are learned through experience (Andrews, Veldre, and Clarke, 2020). This is an important skill for reading, but only one aspect of many. Author recognition is thought to index time spent reading more generally (Acheson, Wells, and MacDonald, 2008). However, this measure likely does not fully capture all types of reading, for instance time spent reading in many online contexts or as subtitles in videos. An interesting avenue for future research would be to obtain more measures of reading skill and experience and assess how they are associated with visual crowding. Another interesting test of the effects of reading on crowding would be to examine letter crowding for readers of languages with different attentional patterns than English. On the basis of the current findings, we would predict that experience reading right-to-left scripts (e.g. Arabic or Hebrew) might create a bias toward the left visual field instead of the right visual field. Similarly, experience reading more complex logographic scripts (e.g., Chinese)—where there is more information concentrated within a fixation—may change the nature of the eccentricity scaling for letter crowding.

We found smaller crowding thresholds in the right visual field than in the left visual field in all three experiments. This effect was similar in magnitude for upright letters, inverted letters, and Gabor patches, so it is not clear whether this effect is related to reading experience. It could be an experience-independent bias that is related to intrinsic differences between the left and right hemispheres (e.g. Michael and Ojeda, 2005), although interestingly, crowded letters in the left visual field benefit more from attentional cues than crowded letters in the right visual field (Ramamurthy et al., 2021). Another factor is that words are ultimately processed in the left hemisphere, so letters in the right hemifield may receive privileged processing (Yeatman and White, 2021). However, this would not explain the fact that a right hemifield advantage was also observed for Gabor patches. Alternatively, experience with reading left-to-right scripts could produce a generalized attentional bias for all stimulus types. This alternative is consistent with the finding from Study 2 that the right hemifield advantage was greater in people with higher scores on both the spelling test and the author recognition test. Previous research does show that increased spelling ability and reading comprehension are associated with larger visual spans in the right visual field for English readers (Veldre and Andrews, 2014). One possibility is that the development of visual span and parafoveal processing during reading in children (Kwon, Legge, and Dubbels, 2007) occurs in tandem with the development of letter-specific mechanisms to reduce crowding. The mechanism for this biased tuning may also be related to the model of crowding put forth by Nandy and Tjan (2012), where the tuning of lateral interaction zones outside of the fovea is learned though spatial attention but biased by eye movements. Given that English reading puts particular constraints on spatial attention and eye movements that are focused in the right visual field, it is possible that this learning process is one source of the general bias of lower crowding distances in the right visual field.

The large sample size of 250 observers in Study 2—to our knowledge, the largest sample in any crowding study published to date—made it possible to demonstrate that crowding varies considerably across individuals and is associated with high-level factors such as spelling ability and working memory capacity. The relationship was relatively strong for working memory capacity, with the model suggesting an overall decrease in crowding threshold of about .034φ for each increase of one standard deviation of working memory capacity. This is consistent with electrophysiological evidence that letter identification recruits working memory under conditions of close spacing (Bacigalupo and Luck, 2015), and suggests that the intersection of visual working memory and crowding may be a fruitful area for future research. A recent letter crowding study with 50 observers (Kurzawski et al., 2023) also found high variability across individuals, and lower crowding thresholds in the right hemifield, similar to the present study.

Such results suggest that to explain the high variability in crowding distances across observers and across the visual field, observer-level characteristics must be taken into account. Our results suggest that individual differences related to reading experience and visual working memory explain some of this variability.

The present evidence for experience-driven changes in a fundamental aspect of visual processing extends previous evidence for a high degree of plasticity in the human visual system. Perceptual learning within the visual system has been observed in a wide array of laboratory tasks (Li, 2016), over longer term experience (Chopin, Bediou, and Bavelier 2019), and in instances of disordered vision (Castaldi, Lunghi and Morrone, 2020). Other studies have shown effects of shorter-term experience on crowding for letters (Chung, 2007; Hussain et al., 2012) and on crowding for other learned, specialized visual stimulus classes such as musical notes (Wong & Wong, 2016). The present findings indicate that crowding is also influenced by a lifetime of reading experience. One plausible mechanism for the effect of reading experience on crowding would be a modification of the receptive field properties of neurons involved in letter recognition such that they integrate letter feature information at smaller scales to reduce interference and optimize the parallel recognition of letters (Chanceaux & Grainger, 2012). A related possibility is that massive experience with letters allows readers to represent letter information at a lower level in the visual hierarchy, where receptive fields are smaller, which fits with the current finding that at least some of the experience-related changes in letter processing are spatiotopic.

Chapter 3: Evidence of early visual tuning for letters revealed by ERP decoding

Reading is a culturally invented behavior that is now essential for individuals of most societies. It is highly practiced, to the point of feeling effortless for many readers, but it is a demanding task for the visual system. Reading is a complex but also highly structured behavior: readers must perform quick and precise eye movements to efficiently sample sentences, recognize words from sets of highly similar and crowded letters, and integrate this visual information with phonological, semantic and syntactic structures. The difficulty and specificity of this behavior requires significant tuning in perceptual systems to produce an efficient reading system. Indeed, learning to read produces widespread changes in the visual system, especially in left ventral occipitotemporal cortex where literacy is thought to cause areas of cortex to be repurposed to recognize written language (Dehaene and Cohen, 2007). Most notably, this leads to the formation of the so-called visual word-form area (VWFA), a region that is critical for visual word recognition (see Dehaene et al., 2010; Dehaene et al., 2015). Various studies have shown that this area responds more strongly to written words than to scrambled words or non-lexical objects (Szwed et al., 2011), but is also tuned to sublexical features of words including single letters (Dehaene et al., 2004) and bigrams (Binder et al., 2006). Although there are fewer studies focused on the earliest visual areas, some studies indicate that tuning for known words exists as early as V1/V2 in French readers (Szwed et al., 2011; Szwed et al., 2014) and as early as V3/V4 for Mandarin readers (Szwed et al., 2014). However, it is unclear whether effects in early visual areas or even the VWFA are representative of selective, feedforward tuning, or represent later feedback (e.g., prediction error signals) from higher level language networks (see Price & Devlin, 2011).

The temporal resolution of electrophysiological measures provides another avenue to uncover tuning in early visual processing due to reading experience. Item-level analysis of ERP responses to visual words have shown that lexical frequency—a measure of how often a word is used in its language—shows relatively early effects starting around 150 ms (Hauk et al., 2006; Chen et al., 2015; Winsler et al, 2023). This indicates some degree of early visual tuning because more frequent words will have been read more often, leading to more tuning in the visual system. However the clearest findings indicating early visual tuning to orthographic stimuli have been from studies of the visual N1/N170 ERP component[1] in response to orthographic stimuli. The N170 was first observed in studies of face perception, as a larger negativity around 170 ms to faces compared to other objects such as cars, particularly posterior electrodes over the right hemisphere (Bentin et al., 1996). A subsequent study by Bentin and colleagues showed that orthographic stimuli (e.g. words) generate larger N170s than non-orthographic stimuli (e.g. symbol strings), but that this N170 effect was largest in the left hemisphere (Bentin et al., 1999). This hemispheric difference between the right-lateralized face N170 and the left-lateralized word N170 is consistent with the lateralization of locations cortical regions involved in processing these stimuli classes; for instance the fusiform face area in right IT cortex (Kanwisher & Yovel, 2006) and the VWFA in left IT cortex (Cohen and Dehaene, 2004).

In general, the N170 is thought to be an index of specialized visual processing for behaviorally important and highly-learned visual categories such as faces and words. One key piece of evidence supporting the idea that the orthographic N170 results from tuning due to experience is that orthographic N170 effects emerge and become more left-lateralized as children

---

[1] There is inherent ambiguity in assigning effects to specific ERP components. Our view is that the N170 is a special case of visual N1-type responses to highly learned stimuli. We therefore will describe ERP responses to orthographic stimuli during the N1 timeframe as orthographic N170s, but acknowledge that it is not clear whether the same component is being described in all of these studies

develop and become better readers (Maurer et al., 2006; Maurer & McCandliss 2007; Uno, Kasai, & Seki, 2024). One informative study using a sample of literate and illiterate adults found that reading ability was associated with larger ERP responses during an N170 window to multiple categories of visual stimuli, but the left-lateralization of responses to letter strings was particularly correlated with reading ability (Pegado et al., 2014).

In alphabetic scripts, words are made up of combinations of letters, and it is generally accepted that the recognition of letters is the starting place of the visual word recognition process (Grainger, 2018). Hence, measuring perceptual processing during the recognition of single letters may be the earliest point at which orthographic tuning may be observable. Some studies have revealed that like words, single letters elicit larger N170s compared pseudo-letters (Stevens et al., 2013; Toussi et al, 2023), although another study did not observe an N170 difference for single letters and symbols presented in the parafovea, but rather observed a difference in N170s to crowded arrays of 3 letters or symbols (Winsler, Grainger, and Holcomb, 2022). A common manipulation in N170 studies is the inversion of stimuli, which delays the latency and increases the amplitude of N170 responses for both faces (Rossion et al., 2000) and words (Rossion et al., 2003) relative to their upright versions. This is a particularly good manipulation for examining effects of experience on visual processing of orthographic stimuli because it breaks familiarity, while fully controlling for many low level features. To our knowledge, there are no published N170 studies with single inverted letters, although one study did compare ERPs to single letters and mirror reversed letters and found larger P1 and N170 responses to reversed letters in adults but not children (Blackburne et al., 2014).

Research on the orthographic N170 has provided a lower bound on the onset of visual processing that is tuned by experience. That is, the finding that ERP amplitudes differ between upright and inverted words by approximately 150 ms indicates that the brain is differentially processing items from these categories by this time. Given that the difference between what constitutes an upright versus inverted word is defined by the experience of learning to read, this indicates that experience-based tuning is affecting visual processing by at least as early as 150 ms after stimulus onset. However, comparing familiar stimuli (e.g., upright letters) and unfamiliar stimuli (e.g., inverted letters) using ERPs that are averaged across many different exemplars may be a blunt tool to study experience-based tuning in visual cortex because it is unclear exactly what a larger ERP indicates in terms of the underlying information processing. Like the arguments against specialized orthographic-specific tuning in the VWFA (Price & Devlin, 2011), orthographic N170 effects may not only represent feedforward processing but rather, feedback prediction errors, increased attentional processes, or recruitment of additional networks which may or may not be tuned for orthographic processing. That is, it is unclear whether greater amplitude indicates improved tuning or some less specific process. Indeed, the fact that the P1 and N1 are larger for mirror-reversed letters than for normal letters, and larger for nonwords than for words, might suggest better tuning for unfamiliar stimuli, which would be the opposite of the hypothesis that experience leads to better tuning.

Another approach is to use multivariate decoding techniques that can quantify how much information about word identity or letter identity is present in the neural signal, which would make it possible to determine whether more information is being extracted about the familiar category (e.g., upright words) than about the unfamiliar category (e.g., inverted words). One study using MEG data showed that letters and numbers can be dissociated from false fonts by around 100 ms (Nara et al., 2023). Another study showed Chinese characters can be distinguished from random stroke combinations also by about 100 ms (Lui et al., 2021). Results like these push the possible time course of experience-tuned orthographic representation activation even earlier. However, these category-level decoding results still do not assess the amount of information being extracted about the identities of the items from the familiar and unfamiliar categories. One study did show that individual short words could be classified and image-reconstructed from EEG about 100 ms after word onset (Ling et al., 2019); however, this study did not include stimuli from an unfamiliar category, so it cannot be used to test hypotheses about the effects of experience on neural tuning.

The current study aims to use multivariate ERP decoding analyses in addition to univariate ERP analyses to uncover the earliest point in the ERP time course that is affected by reading experience and to determine whether reading experience changes the tuning of early visual processing. To do this, we developed a paradigm in which a very large number of trials was obtained for a set of 6 letters, with each letter presented upright on some trials and inverted on other trials. This gave us the signal-to-noise ratio to detect small but early changes in ERP amplitudes and to decode the identities of the individual upright letters and the identities of the individual inverted letters. We also obtained data from a larger-than-typical sample of participants to further increase statistical power. We then conducted traditional univariate analyses to compare the amplitudes of the signals generated by the set of upright letters versus the set of inverted letters. We followed this with multivariate pattern analyses to compare decoding accuracy for distinguishing among the upright letters with decoding accuracy for distinguishing among the inverted letters.

Following from previous results showing larger N170s to inverted words (e.g. Rossion et al., 2003), we predicted that the N170 would be larger for inverted letters than for upright letters. If inverted orthographic stimuli produce an increased N170 because of enhanced processing (e.g., due to longer processing or larger prediction errors), then decoding accuracy should be greater for inverted letters than for upright letters beginning in the N170 time range (e.g., beginning approximately 150 ms after stimulus onset). In contrast, if the larger N170 reflects enhanced neural tuning for upright letters than for inverted letters, then decoding accuracy should be greater for upright letters than for inverted letters. A third possibility is that decoding accuracy will be greater for upright than inverted letters initially, reflecting enhanced tuning, followed by greater decoding accuracy for inverted than upright letters, reflecting extended processing for the inverted letters (potentially as a result of error signals feeding back from higher-level systems). This would fit with the evidence of N170 latency shifts, with inverted stimuli (e.g. Rossion et al., 2003) generating slightly later N170s than their upright counterparts.

Following from previous results showing larger P1s to reversed letters (Blackburne et al., 2014), we further predicted that P1 amplitude would be greater for inverted letters than for upright letters. Just as with the N170, we will use letter decoding accuracy to determine the nature of this effect. If early visual processes are better tuned to discriminate among upright letters than among inverted letters, then accuracy for decoding letter identity should be greater for upright letters than for inverted letters. If, in contrast, inverted letters attract greater processing resources very rapidly, then decoding accuracy should be greater for inverted than upright letters.

Further, we presented the stimuli at three different locations—fovea, left parafovea, and right parafovea—to determine whether the patterns of ERP amplitudes and decoding accuracy differ across retinal locations. Because words are letters are primarily read in the fovea and in the right parafovea (in the Latin script – see Schotter et al., 2012), any early visual tuning for letters may be retinotopic (e.g. Winsler & Luck, 2024). If this is the case, initial decoding differences between upright and inverted letters may be larger in the fovea and right visual field compared to the left visual field.

Method

Participants

Data from 30 participants were used in all analyses. This sample size was chosen prior to data collection to have high power to detect subtle early ERP effects, and has about twice the number of participants than similar studies (e.g. Rossion et al., 2003; Blackburne et al., 2014: Stevens et al., 2013). All participants were students enrolled at the University of California, Davis. All participants were native and current readers of a Latin script (typically English or Spanish), native being defined as learned as their first written language or concurrently with their first written language. All participants reported having normal or corrected-to-normal vision. Some participants were rejected on the basis of poor behavioral or EEG data (discussed below), and these participants were simply replaced until we reached our target N of 30.

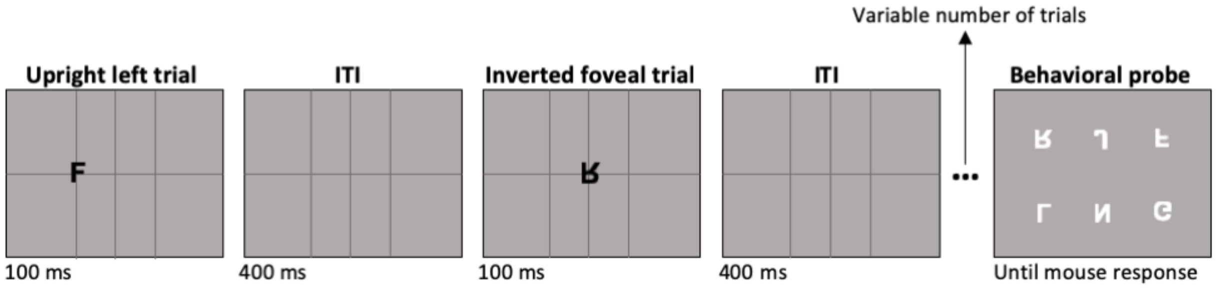

Figure 6. Trial schematic. After a random number of letter stimuli, a behavioral probe array appeared, and the participant was required to click on the letter that matched the identity of the

immediately preceding letter. The letters in the behavioral probe array were upright if the preceding letter was upright, and they were inverted if the preceding letter was inverted.

Materials and procedure

Stimulus presentation was controlled using PsychoPy (Pierce, 2007). Stimuli were presented on a HP ZR2440w LCD display with a gray background (26.8 cd/m^2) at a nominal viewing distance of 100 cm.

Over the course of the experiment, participants viewed single letters appearing at 3 locations: centered at fixation or 1.5° of visual angle to the right or left of fixation. The eccentricity of 1.5° for the left and right locations was chosen such that the right location fell roughly within the asymmetrical perceptual span of Latin script readers (see Rayner et al, 2010). The letters were a set of six black capital letters (F, G, J, L, N, R) drawn in the Arial typeface with a height subtending 1.0° of visual angle. These letters were either in their familiar upright orientation, or inverted (i.e., flipped vertically). The set of letters were selected such that their upright and inverted forms were different from each other. In total there were six conditions, three Location (left, foveal, right) by two Familiarity (upright or inverted). In each of these six conditions, each of the six letters were presented 120 times, leading to a total number of 4320 trials. These trials were randomly shuffled within twelve blocks of 360 trials.

The stimuli and task are illustrated in Figure 6. Each trial began with the participant fixating at the center of the screen, which was marked by the intersection of horizontal and vertical thin dark-gray lines that extended to the edge of the screen. There were also two vertical lines at the left and right locations, so that all stimuli were presented on top of intersecting horizontal and vertical lines. After 400 ms of this ITI screen, an upright or inverted letter was presented for 100 ms at one of the three locations on top of the dark-gray lines. Participants were instructed to maintain fixation at the central location. The task for participants was to recognize each stimulus as it appeared so that they could report the letter identity if it was followed by a probe array, which occurred after a random 0.5% of trials. The probe array occurred after the ITI screen, and it contained the 6 possible stimuli from the set of letters, randomly distributed in a 3x2 grid. The ordering of the letters in the grid varied randomly from array to array so that participants would not learn to associate a letter identity with a specific location. The letters in the behavioral probe array were upright if the preceding letter was upright, and they were inverted if the preceding letter was inverted. Subjects used a mouse to click on the stimulus that matched the target from that trial. After response, the background lines reappeared, and a shrinking circle at the center of the screen drew the participant's attention back to the fixation point for 1200 ms, followed by the ITI screen and then the next trial. Each participant was probed 18 times each block, for a total of 216 probes. These were shuffled such that there were no fixed number of how many trials passed before being probed, but on average it would be 20 trials (corresponding to 10 seconds), with a maximum of 39 trials between probes.

In this task, chance was 1/6 (16.7%). As a quality control measure, participants were excluded from all analyses if their accuracy in selecting the probe letter was less than 4/6 (66.7%,corresponding to an accuracy of less than 4 times chance). Six subjects were replaced for this reason.

Behavioral data analysis

Behavioral performance on probe trials were modeled using logistic linear mixed effect regression as implemented by the lme4 R package (Bates, et al., 2014). The response variable was whether each of the 216 probe trials was answered correctly (1) or incorrectly (0). The fixed effect structure of the model included effects for Familiarity (upright vs. inverted letters), Location (foveal, left, right), and the interaction between the two. These variables were dummy coded, with upright foveal letters serving as the reference group. The random effect structure included random intercepts for Stimuli (each of the 12 stimuli), random intercepts for Subject, and by-subject random slopes for Familiarity and Location. Effect significance was tested with Wald chi-square tests using Satterthwaite's method through the lmerTest package (Kuznetsova, Brockhoff, and Christensen, 2017). All tests used an alpha of .05.

EEG Recording and Processing

EEG data was recorded from 32 electrodes using a Brain products actiCHamp system. This included electrodes at 27 scalp sites (FP1, FP2, F3, F4, F7, F8, C3, C4, P3, P4, P5, P6, P7, P8, P9, P10, PO3, PO4, PO7, PO8, O1, O2, Fz, Cz, Pz, POz, Oz), electrodes on the left and right mastoids, an electrode placed below the right eye, and electrodes placed lateral to the left and right eyes. Impedances were kept below 25 kΩ. EEG was recorded at 500 Hz using a cascaded integrator-comb anti-aliasing filter with a half power cut of at 130 Hz. Offline, data were processed using the Matlab packages EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). Event codes were shifted forward 28 ms to account for stimulus monitor presentation delay. The data were then down sampled to 250 Hz, and filtered using a noncausal Butterworth bandpass filter (0.1 – 30 Hz) with a 12 dB/octave roll off.

To aid in eye-movement artifact rejection, bipolar HEOG channels (a subtraction of horizontal left and right eye channels) and VEOG channels (a subtraction of FP2 and lower eye channels) were created. Periods of extreme EEG voltages were removed from the continuous data, then the data from the original channels were decomposed using independent component analysis. Components determined to be blinks were removed from the data, and then the data were re-constructed. The EEG data was then re-referenced to the average of all scalp electrodes and was epoched from -100 to 400 ms relative to stimulus onset. Then artifact rejection was performed on all channels, rejecting trials with extreme values in any channel. Trials containing artifacts were excluded from both the averaged ERP analysis and the decoding analysis. Participants who had fewer than 90 (out of 120) remaining trials for any item-by-condition combination (e.g., upright "F" trials in the foveal condition) were excluded from further analysis to ensure sufficient trial numbers for the decoding analysis. This reflects an a priori policy of excluding participants for whom more than 25% of trials are rejected (Luck, 2014), but updated for decoding studies to apply to every item-by-condition combination. We replaced 1 participant for this reason.

Univariate ERP analysis

Averaged ERPs were created over the epoch of -100 to 400 ms, using the pre-stimulus period as a baseline. These averages collapsed across the individual letter identities, separately for upright and inverted letters. Following from previous N170 studies, for the univariate analyses we focused only on posterior-occipital electrodes and as recommended by Zhang and Kappenman (2024). These channels were averaged into one cluster for the left (P7, P9, PO7, O1) and one for the right (P8, P10, PO8, O2).

The ERPs for upright and inverted letters were compared at each of the locations separately, due to the possible timing differences according to contralateral and ipsilateral presentation. Statistical analyses used a mass univariate analysis (Groppe, Urbach, & Kutas, 2011), conducted on the difference wave between upright and inverted letters (inverted minus upright). In this analysis, two-tailed, one-sample t-tests are performed against 0 at each time point (each 4 ms) in the analysis window, and FDR corrected using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). Our time window for both the averaged ERP analysis and the decoding analysis started at 64 ms and extended to 300 ms. The start time of this window was chosen to match when location information was first able to be decoded (explained below) and represents the earliest possible time we could detect a difference between upright and inverted letters. The end time of the window was chosen to include the full extent of possible N170 effects.

ERP decoding analysis

The main focus of the decoding analyses was to decode stimulus identity separately for upright and inverted letters, and separately at each of the 3 locations (Left, Foveal, Right). Decoding analyses were run separately for each participant at each 4 ms time point using leave-one-out 3-fold cross-validated linear support vector machines (see Carrasco et al., 2024; Zhang et al., 2024). The decoders were trained and tested using the amplitude for each scalp electrode at that time point. Trials with artifacts were excluded. Under this approach, for each participant and stimulus class (i.e., letter), trials were randomly subsampled and averaged together to create 3 averages containing the same number of trials. For analyses comparing upright and inverted letters, the number of trials per average was balanced such that the same number of trials was used for decoding each upright or inverted letter. The decoding was cross-validated such that on each iteration, each average was used to test the decoder once and train the decoder twice. The decoding accuracy for each of these folds was averaged to create the decoding accuracy for that iteration. This process was repeated for each time point in the ERP, which was every 4 ms from -

100 ms to 400 ms. Then the whole process was repeated with a new random assignment of trials to averages for 100 iterations, and accuracy was averaged across folds and iterations to yield the decoding accuracy for that subject and time point. All support vector machines used the scikit-learn "SVC" implementation (based of the LIBSVM library; Chang and Lin, 2011) with a one vs. one classification scheme and the C parameter set to 1.

This analysis was run separately for each location and familiarity condition (upright or inverted). Additionally, to examine possible differences in the information present between the left and right hemispheres for letter decoding, we conducted the same decoding analysis using the data from only the left channels (FP1, F3, F7, C3, P3, P5, P7, P9, PO3, PO7, O1) or only the right electrode channels (FP2, F4, F8, C4, P4, P6, P8, P10, PO4, PO8, O2).

Paralleling the univariate analyses, to test whether decoding accuracies were above chance, the accuracy for each time point in the analysis window was tested against chance (1/6 = .125) using one-sample t-tests. These tests were one-tailed because only noise can produce below-chance values in this specific analysis. To test whether decoding accuracy was different between upright and inverted letters, the decoding accuracy for each condition at each time point was compared using paired-sample two-tailed t-tests. All p-values were FDR corrected using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). As with the univariate analysis, the analysis window was each time point between 64 ms – 300 ms. The start time of this window was chosen as the first time point in which the location of the stimuli could be decoded significantly above chance, which should represent the absolute earliest point at which the identity of the stimuli could be decodable. This was determined using the same decoding procedure described above, except all trials were used to decode the location of the stimuli (left, foveal, right) regardless of the identity of the stimuli. Significance of location decoding accuracy was tested using one-sample, one-tailed t tests on all time points between 0 and 300 ms, and the time point that was first significant was 64 ms. The location decoding results are shown in Figure 7.

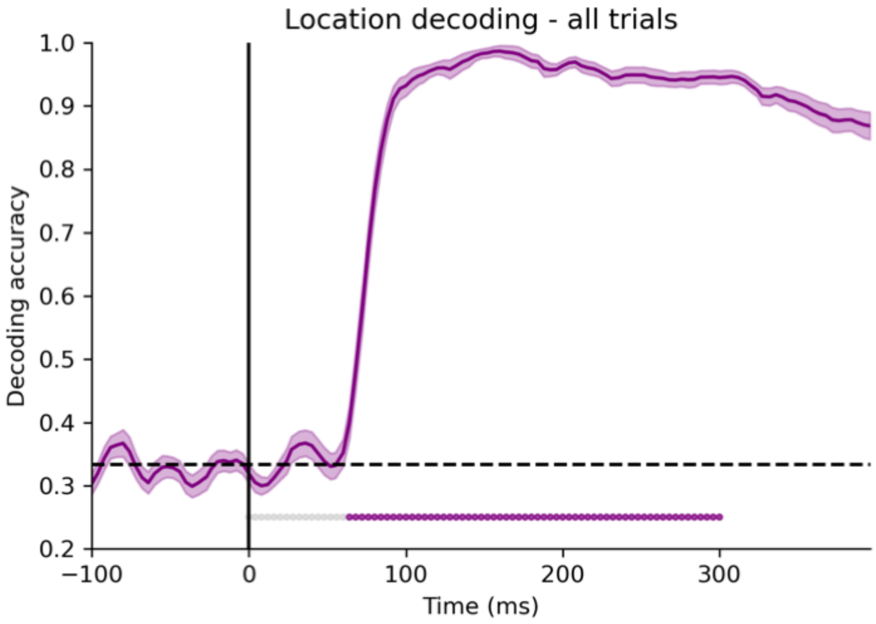

Figure 7. Location decoding results. Stimulus location (Foveal, Right or Left) could be decoded significantly better than chance starting at 64 ms, representing an estimate of the earliest point when information reached visual cortex. Significance was tested from 0 to 300 ms, FDR corrected. Significance for each time point is marked with filled-in dots at the bottom of the plot.

Results

Behavioral probe accuracy results

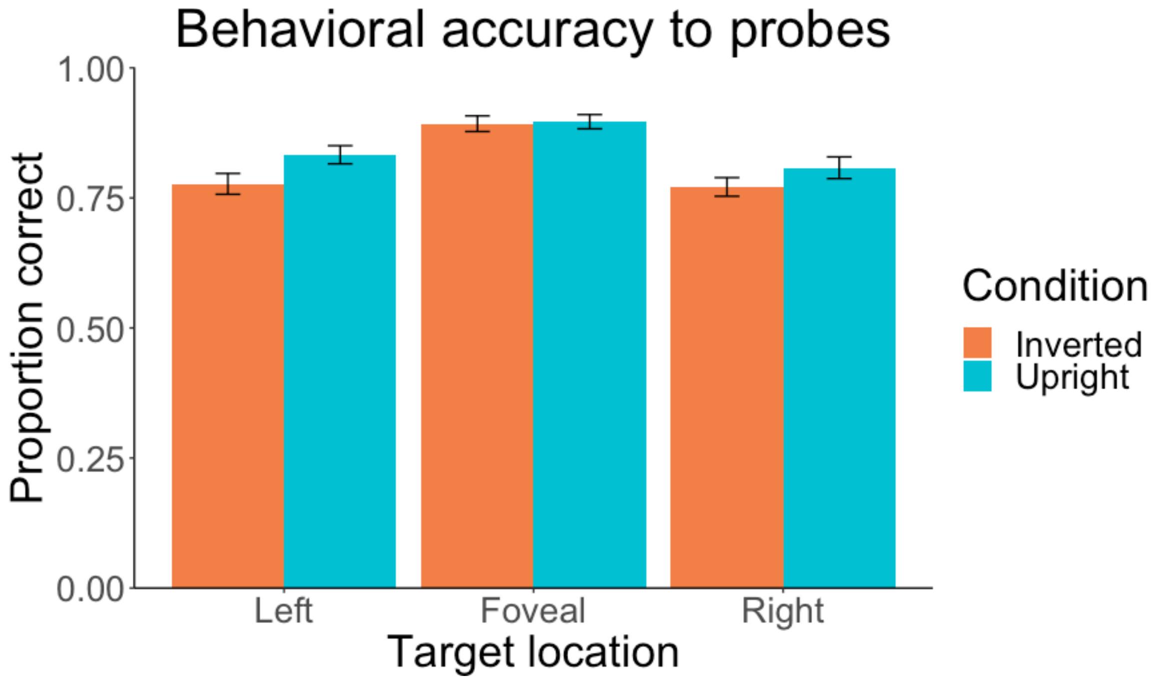

Figure 8. Proportion correct for behavioral probes that were randomly distributed throughout the experiment. Probes asked participants to report the immediately preceding stimulus and chance was 1/6. Error bars represent +/- 1 SEM

Behavioral results are summarized in Figure 8. Accuracy on probe trials showed a significant effect of Location, with higher accuracy for foveal probes compared to left or right probes, X^2 (2, $N = 30$) = 37.62, $p < .001$. Condition showed a marginally significant effect, with upright letters showing higher accuracy than inverted letters, X^2 (1, $N = 30$) = 3.19, $p = .074$. The interaction between Condition and Location was not significant (p = .306). A follow up model with an identical structure, but fit with only the left and right location probes, showed a significant effect of Condition with upright letters showing higher accuracy than inverted letters, X^2 (1, $N = 30$) = 4.59, $p = .032$. In this model, there was no significant difference between the left and right locations (p = .271) and again no significant interaction between Condition and Location (p = .596).

Univariate ERP results

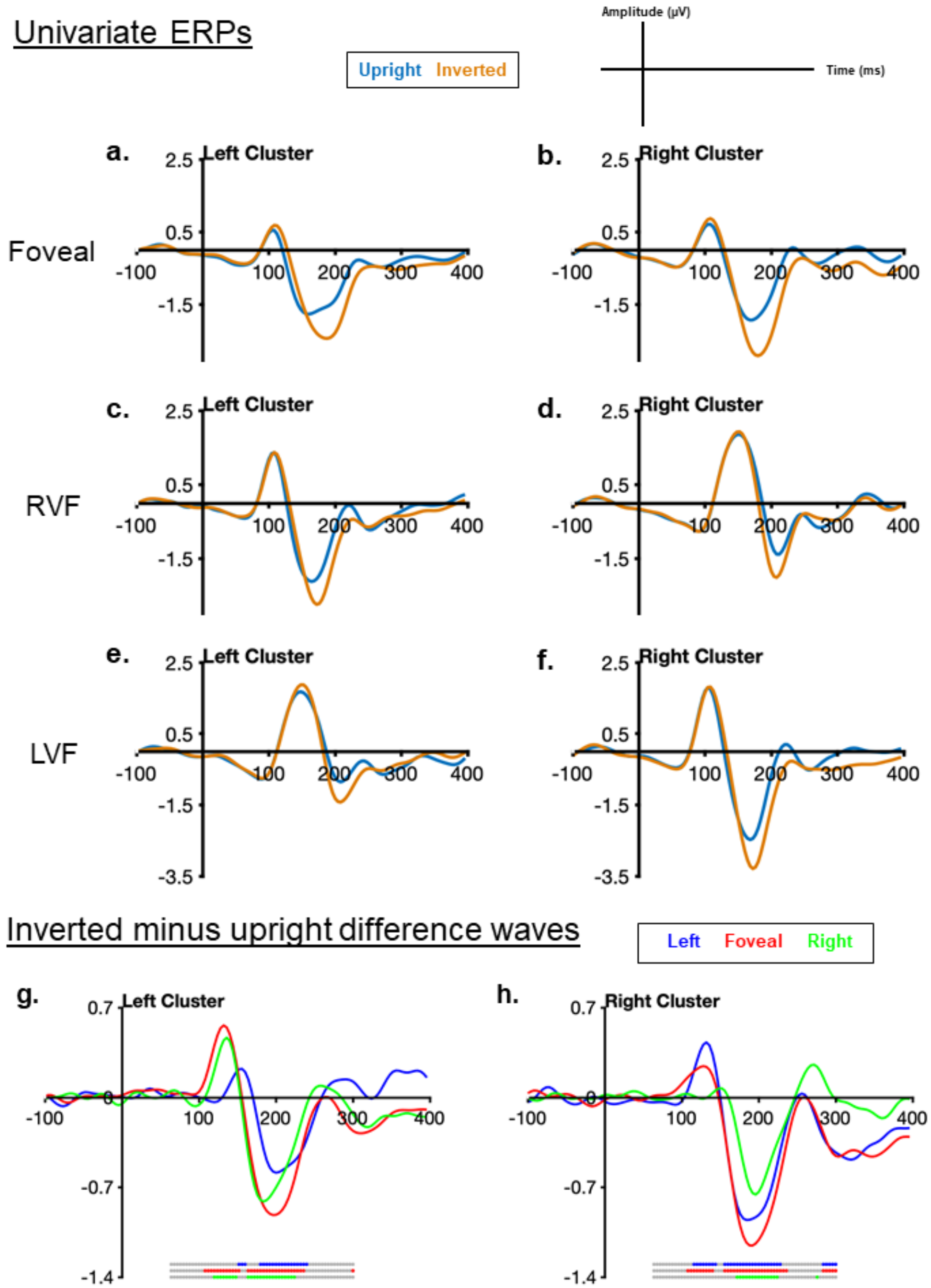

Figure 9. Univariate ERPs in posterior left and right electrode clusters for upright versus inverted letters presented in the fovea (**a,b**), in the right visual field (**c,d**), and in the left visual

field (**e,f**). Inverted minus upright difference waves for each location in the left (**g**) and right (**h**) electrode clusters. Colored dots beneath the difference waves mark time points when the difference between inverted and upright letters was significantly different from 0 (FDR corrected). Grey dots indicate time points that were tested (64 – 300 ms) but were not significant.

We begin by presenting the univariate analyses of averaged ERP amplitudes, collapsed across the individual letter identities. The averaged ERPs are shown in Figure 9a-f, and difference waves (inverted minus upright) are shown in Figure 9g-h. In the parent waveforms, a slightly larger P1 can be seen for inverted letters than for upright letters for the foveal and right stimulus locations, especially in the left hemisphere electrode cluster (Figures 9a and 9c). The N170 was also clearly larger for inverted letters than for upright letters, especially for foveal stimuli (Figures 9a and 9b). These effects can be seen more clearly in the inverted-minus-upright difference waves (Figures 9g and 9h). In the P1 latency range, the difference for foveal stimuli was larger in the left hemisphere than in the right hemisphere, and the difference for lateralized stimuli was larger in the contralateral hemisphere than in the ipsilateral hemisphere. In the N170 latency range, by contrast, the difference for foveal stimuli was larger in the right hemisphere than in the left hemisphere. For lateral stimuli, the N170 difference was larger in the contralateral hemisphere.

For the statistical analysis, we will begin by considering the effects in the P1 latency range. In this time period, the voltage for foveal stimuli was significantly more positive for inverted letters than for upright letters in both the left and right clusters, beginning at 108 ms. For lateralized letters over the contralateral hemisphere, the voltage was also significantly more positive for inverted than for upright letters, but this began slightly later – starting at 116 ms in left channels and at 120 ms in right channels. As expected due to contralateral dominance and the hemispheric transfer time, this effect was smaller and later over the ipsilateral hemisphere, with a P1 effect in the left hemisphere cluster for letters in the left visual field starting at 152 ms and no significant P1 effect in the right hemisphere cluster for letters in the right visual field.

Now we turn to the statistical analysis in the N1 latency range. In this time period, the N170 was significantly larger (more negative) for inverted than for upright stimuli for both the left and right clusters and for all location conditions. For both foveal and contralaterally presented stimuli, this effect became significant in left channels at 164 ms and slightly earlier in right channels at 156 ms. For ipsilaterally presented stimuli, the N170 effect began at 180 ms in the left hemisphere and again slightly earlier in the right hemisphere at 172 ms. These effects on the N170 persisted in all conditions for about 60 ms until roughly 230 ms.

ERP decoding results

Next we used ERPs from all scalp electrodes to decode the exact letter identity, separately for upright and inverted letters at each of the three spatial locations. The decoding accuracy time series are shown in Figure 10.

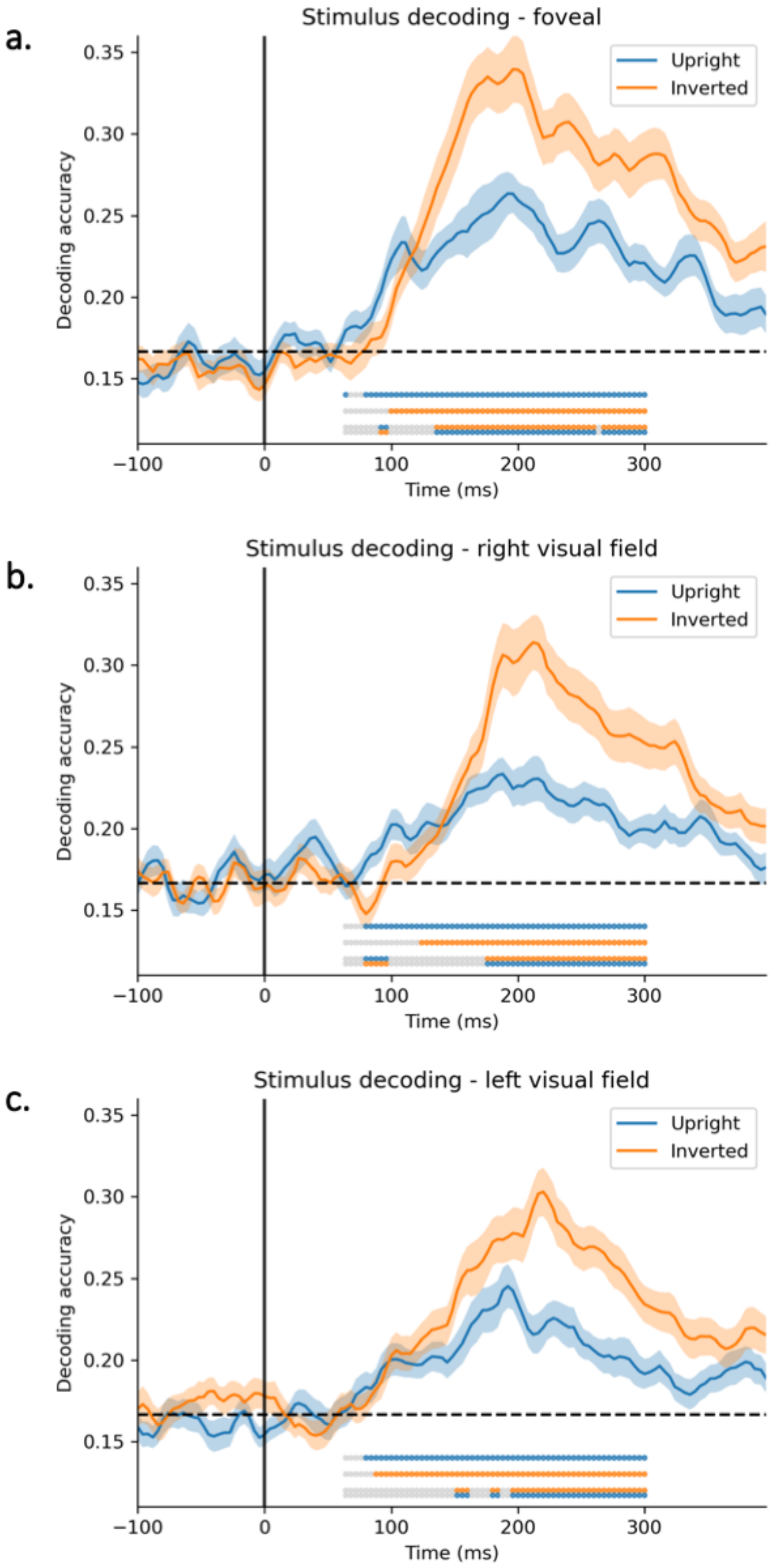
a.
Stimulus decoding - foveal
Upright
Inverted
Decoding accuracy
0.35
0.30
0.25
0.20
0.15
−100
0
100
200
300
Time (ms)
b.
Stimulus decoding - right visual field
Upright
Inverted
Decoding accuracy
0.35
0.30
0.25
0.20
0.15
−100
0
100
200
300
Time (ms)
c.
Stimulus decoding - left visual field
Upright
Inverted
Decoding accuracy
0.35
0.30
0.25
0.20
0.15
−100
0
100
200
300
Time (ms)

Figure 10. Decoding time-courses for upright and inverted letters in the fovea (**a**), right visual field (**b**) and the left visual field (**c**). Colored dots beneath the waves mark time points when decoding accuracy for inverted and upright letters was significantly higher than chance (1/6). Stacked colored dots indicate time points when there was a significant difference in decoding accuracy between upright and inverted letters, with the color on top indicating which condition had the higher decoding. Grey dots indicate time points that were tested (64 – 300 ms) but were not significant. All tests were FDR corrected.

Foveal letters (Figure 10a).

For stimuli presented in the fovea, upright letter identity could be decoded significantly above chance starting at 80 ms, just 16 ms after the location of the stimuli could be decoded (see Figure 7) and well before any significant differences between upright and inverted letters were observed for the averaged ERP amplitudes (see Figure 9). The 64 ms time point also showed significant upright letter decoding, but since it was a single isolated point, we have more confidence that 80 ms is the true onset of upright letter decoding. For inverted letters however, stimulus identity became significantly decodable later, at 100 ms. Moreover, decoding accuracy was significantly greater for upright letters than for inverted letters at the 92 and 96 ms time points. Thus, upright foveal letters could be decoded quite rapidly—well before 100 ms—and they could be decoded more accurately than upright letters before 100 ms.

Decoding accuracy for both upright and inverted letters in the foveal location remained above chance until the final time point of the analysis window at 300 ms. Starting at 136 ms, however, decoding accuracy was significantly greater for inverted letters s than upright letters, and this difference remained significant until the end of the analysis window 300 ms (except for a single time point at 264 ms).

Right visual field letters (Figure 10b).

We observed a similar pattern for stimuli in the right visual field. As with the foveal stimuli, upright letters were decoded above chance starting at 80 ms, whereas decoding accuracy for inverted letters was not significantly above chance until 124 ms — 44 ms later. Moreover, decoding accuracy was significantly greater for upright letters than for inverted letters from 80–96 ms. Decoding for both upright and inverted letters remained above chance until 300 ms. Starting at the 176 ms time point, inverted letters had significantly higher decoding than upright letters, and this effect remained until the final time point in the analysis window.

Left visual field letters (Figure 10c).

In the left visual field, a slightly different pattern of early effects emerged. Both upright letters and inverted letters were still decodable early, starting at 80 ms for upright letters and slightly later at 88 ms for inverted letters. However unlike in the fovea and the right visual field, decoding accuracy was not significantly greater for upright letters than for inverted letters in the early time period. Decoding remained significantly above chance for both upright and inverted letters through 300 ms. As with the other locations, decoding accuracy at later time points was significantly greater for inverted letters than upright letters. This effect became significant at 152 ms and remained significant throughout the analysis window except for two brief periods (164-176 ms and 188-192 ms).

ERP decoding results split by hemisphere

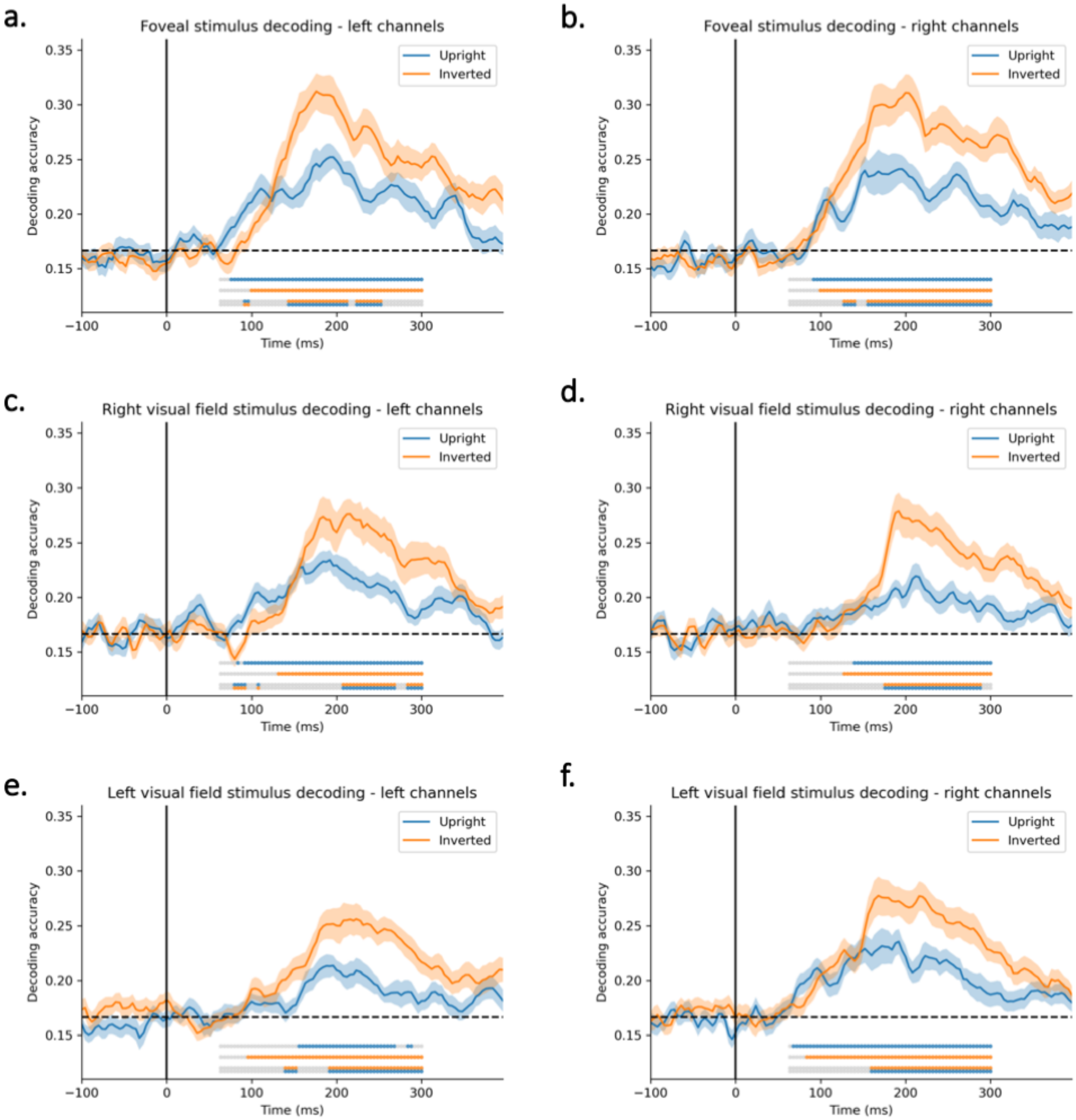

Figure 11. Decoding time-courses in left and right hemisphere channels for upright and inverted letters in the fovea (**a,b**), right visual field (**c,d**) and the left visual field (**e,f**). Colored dots beneath the waves mark time points when decoding accuracy for inverted and upright letters was significantly higher than chance (1/6). Stacked colored dots indicate time points when there was a significant difference in decoding accuracy between upright and inverted letters, with the color

on top indicating which condition had the higher decoding. Grey dots indicate time points that were tested (64 – 300 ms) but were not significant. All tests were FDR corrected.

Foveal letters.

To test for differences between the left and right hemispheres in the neural information about letter identity, we conducted the same decoding analyses separately for left and right hemisphere electrodes. The results are shown in Figure 11.

For the foveal location, left channels (Figure 11a) showed the same pattern of early effects as in the full montage decoding, with significantly decoding starting at 76 ms for upright letters but not until 100 ms for inverted letters, and a significant difference between them at the 92 and 96 ms time points.

Channels over the right hemisphere (Figure 11b) did not show this early effect for foveal letters. In these channels, significant decoding was observed for upright letters at 92 ms and for inverted letters onset at 100 ms, but the decoding accuracy was not significantly great for upright letters than for inverted letters. As was observed for the foveal location when all channels were used, decoding for upright and inverted letters for both right and left channels remained significant until the end of the analysis window at 300 ms, and decoding accuracy eventually became significantly greater for inverted letters han for upright letters. Thus, to the extent that electrodes on a given side reflect the information in the underlying hemisphere, these results indicate that information about letter identity is extracted earlier by the left hemisphere than by the right hemisphere and that the advantage for upright over inverted letters is primarily present in the left hemisphere.

Right visual field letters.

Upright letters presented in the right visual field also showed early above-chance decoding for the left (contralateral) hemisphere electrodes (Figure 11c), with a brief point of significant decoding at 84 ms and then consistently significant decoding starting at 92 ms. By contrast, inverted letters in the right visual field did not show above-chance decoding in these electrodes until 132 ms, and decoding accuracy was significantly greater for upright than inverted letters between 80 and 92 ms and then again at 108 ms (Figure 11c). In the right hemisphere (ipsilateral) channels (Figure 11d), we did not observe these early differences between upright and inverted letters in the left visual field; instead inverted letters were first decodable at 128 ms, upright letters at 140 ms, and significantly greater decoding for inverted than upright letters onsetting at 176 ms. These results provide further evidence that the advantage for upright over inverted letters is primarily present in the left hemisphere.

Left visual field letters.

For letters presented in the left visual field, we observed no early difference in decoding between upright and inverted letters with either the left or right hemisphere electrodes (see Figure 11e-f). In the right (contralateral) channels, upright letters did show early above-chance decoding, starting at 68 ms, whereas above-chance decoding for inverted letters did not start until 84 ms. However there was no significant difference in decoding accuracy between upright and inverted letters in these early time points. Decoding accuracy for both upright and inverted letters remained significant until 300 ms, and inverted letters had significantly higher decoding accuracy than upright letters from 156 ms until the end of the analysis window. Decoding accuracy for inverted letters on the left side was significantly above chance in the left electrodes starting at 96 ms and lasting until the end of the analysis window. Interestingly, upright letters on the left side were not decodable in these channels until 156 ms, and this decoding remained above chance until 272 ms. Decoding accuracy was significantly higher for inverted letters than for upright letters for these channels in a window between 140 ms and 152 ms, and then from 192 ms until the end of the analysis window. These results suggest that there is little or no advantage for upright over inverted letters for stimuli presented in the left visual field.

Discussion

We measured early visual ERP responses to single upright and inverted letters presented in the fovea and in the right and left parafovea. Using classical univariate ERP analyses, we found that upright and inverted letters were differentially processed beginning shortly after 100 ms, with larger P1 and N170 amplitudes for inverted compared to upright letters. Using decoding analyses, we found that during the N170 time frame, decoding accuracy was greater for inverted letters than for upright letters, indicating that the neural signals during this time window contained more information about which letter had been presented for inverted letters than for upright letters. Thus, a larger N170 for inverted letters was accompanied by more information about letter identity for inverted letters.

However, letter identity could be decoded significantly above chance beginning much earlier; in most cases before 100 ms. Decoding accuracy prior to 100 ms was significantly greater for upright letters compared to inverted letters, particularly for letters presented in the fovea and right parafovea. Further, this early difference was found to be primarily driven by electrodes over the left hemisphere. Thus, whereas the P1 wave was larger for inverted letters than for upright letters, the early neural signal contained more information about upright letters than inverted letters. Thus, the most important conclusion from the present study is that the brain extracts more information about upright letters than about inverted letters at an early time point, well before 100 ms, consistent with an effect of familiarity on the tuning of relatively early visual processes.

We found a univariate difference between upright and inverted letters during the P1 time range, with inverted letters generating more positive amplitudes than upright letters. This effect began in both left and right electrode clusters just after 100 ms for foveal stimuli, and onset about 20 ms later for contralateral presented stimuli. The timing of this effect (starting at 108 ms for foveal stimuli) aligns with the findings of Blackbourne et al. (2014), which showed that mirror-reversed letters also elicited larger P1s than normal letters. This timing also aligns with studies which decoded categorical differences between orthographic characters and false fonts (Nara et al., 2023; Lui et al., 2021). One interpretation for the reduced positivity observed for upright letters during the P1 timeframe may have less to do with the P1 per se, but rather with the timing of the N170. Fitting with the evidence that inversion delays the N170 during word or face processing (Rossion et al., 2003), it may be the case that upright letters have smaller P1s because they transition to N170 processing faster than inverted letters. However, comparing aggregated categories whether with univariate ERPs or decoding has more-limited interpretability than decoding the stimulus identity itself.

Importantly, our design allowed us to also decode when stimuli-specific information was present for upright and inverted letters. The identity of upright letters was decodable quite early in the ERP signal. For all three locations (fovea, left and right), upright letters were reliably decoded above chance starting at 80 ms. This is just 16 ms after the location of stimuli was able to be to be decoded, which represents the earliest point in the ERP time course at which we can safely assume visual information has reached visual cortex. This decoding of letter identity was earlier than other published results showing that the identity of 3-letter words can be decoded starting at about 100 ms (Ling et al., 2019). Naturally, this could be because our study used isolated letters, which are necessarily more basic than words, as words are made up of letters. Additionally, the current study had very high statistical power as a result of the relatively large number of trials and participants, which may have allowed the decoding of subtle differences in the multivariate patterns of activity that dissociate between letter identities at very early stages in the ERP time course.

Perhaps most interestingly, the onset of identity decoding for upright letters preceded that of inverted letters. Inverted letters could be decoded starting at 100 ms in the fovea, at 132 ms in the right visual field, and at 88 ms in the left visual field. This difference of higher decoding accuracy for upright letters compared to inverted letters was significant only for stimuli in the fovea, starting at 92 ms, or in the right visual field, starting at 80 ms. Because of the high degree of visual similarity between the upright and inverted letters, it is unlikely that this difference in decoding onset is due to anything other than the participant's experience recognizing upright letters. That is, upright and inverted letters and differed only by a transformation (inversion) that was identically applied to all letters in the inverted condition. This means that the relative degree of similarity or dissimilarity with regard to any basic visual feature (e.g. luminance, spatial frequency, orientation, perimetric complexity) between members of the set of inverted letters was identical to those in the set of upright letters. The difference between the two sets lies in the fact that conjunctions among these features specifically for upright letters better matches visual letter representations that have been extensively learned through reading. Hence, we interpret the difference in early decoding between upright and inverted letters as indicative of tuning in visual cortex for letters due to experience.

The finding that this pattern was only present for stimuli in the fovea or right visual field is consistent with the interpretation that this difference in early decoding is due to reading experience. While reading relies heavily on the fovea, readers also extensively sample orthographic information from upcoming words in the parafovea (Schotter et al., 2012). For readers of Latin scripts which are read from left-to-right, this span ranges from about 15 character spaces to the right of fixation compared to about 4 spaces to the left of fixation (see Rayner, Slattery and Belanger, 2010). Hence, this may spatiotopically bias the tuning of letter representations towards the right visual field, potentially contributing to the difference we observe in early decoding between the left and right locations. However another clue as to the nature of this tuning may be found in the decoding analyses split by channels in over the left and right hemispheres. Here, we only observed significant differences in early decoding between upright and inverted letters in the left-hemisphere channels. While this may also be related to reading direction—as information from the right visual field first projects to left visual cortex—this may also reflect the left-lateralized nature of language processing. Indeed, the left-lateralization of N170 is thought to result from processing in left-lateralized language regions like the VWFA, although there is limited direct evidence for this (see Blackbourne et al., 2014 for one example). It may be possible that fast feedforward signals during the early visual processing of upright letters begins to activate higher level letter representations in left ventral-occipital areas like VWFA as early as 80 ms. This would be extremely fast, but not unprecedented – for instance intracranial recordings in monkeys shows feedforward information can flow from V1 to IT cortex in about 25 ms (Schroeder et al., 1998). Unfortunately, based only on the hemisphere differences in the current results, we cannot dissociate between early visual tuning for letters within left occipital cortex due to reading direction, tuning for letters causing early feedforward activation of higher level letter representations, or a mixture of both. Future research using a similar design but with readers of a right-to-left language such as Hebrew or Arabic would help to disambiguate the contributions of reading direction and language lateralization to visual tuning for letters.

After the early decoding and univariate P1 effects, we also observed a clear univariate N170 effect. These results followed from a clear prediction based on prior N170 studies. That is, inverted faces and words generate larger N170s than their upright counterparts, and our study shows that this is the case for single letters as well. While this was not particularly surprising, it was important to verify since the processing demands for letters versus full words are different. That is, although word recognition involves letter processing, the recognition of single letters does not necessarily involve word processing. Our results add to previous results showing that letters produce larger N170s than pseudo-letters (Stevens et al., 2013; Toussi et al, 2023) and reversed letters produce larger N170s than letters (Blackburne et al., 2014), and suggest that the generation of an orthographic N170 effect involves processing at the level of individual letters. This fits with a study using fMRI and MEG that suggests letter selectivity in a region just anterior to the VWFA at about 160 ms (Thesen et al., 2012). Future studies using isolated letters, non-word letter strings, and words may be useful for detangling the contributions of letter-level and word-level processing to the generation of the orthographic N170.

During and after the N170, both upright and inverted letters were able to be decoded well above chance. However, during this time period decoding accuracy was much higher for inverted letters than upright letters. The timing of this effect had some variability, beginning in the all-channel analysis starting at 136 ms, 152 ms, and 176 ms for stimuli in the fovea, left, and right visual field respectively. However it did eventually emerge regardless of the location the stimulus was presented or which hemisphere electrodes were used to do the decoding. The timing of this effect largely paralleled the onset of the univariate N170 effects, and the direction of the two effects matched in the sense that inverted letters had both larger N170s and higher decoding accuracy. This has important implications for understanding the generation of univariate orthographic N170 effects. Specifically, this suggests that larger amplitude N170s to inverted letters corresponds to more stimuli-specific information present in the EEG signal. This indicates that the generation of larger N170s is not because of the recruitment of neurons auxiliary to the stimulus representation (e.g. general attentional processes or generic error signals), but rather because of increased processing that results in a higher amount of discriminable stimulus information. The finding that decoding accuracy was higher for the less-familiar stimulus category (inverted letters) may indicate that representing inverted letters requires longer or more-in-depth processing which in turn causes the activation of more stimulus-specific information that is relayed in the EEG signal. Relatedly, predictive coding frameworks would predict that less-familiar stimuli could produce larger prediction errors, which—if specific enough to encode letter or letter feature information—may also contribute more decodable signals. Although we only tested upright and inverted letters, given that the pattern of univariate N170 inversion effects we observed for single letters was similar to what others have observed with words and faces (Rossion et al., 2003), it would be reasonable to expect that other stimulus classes would show a similar correspondence between larger N170s and higher decoding accuracy.

In conclusion, this study used classical univariate ERP analyses paired with multivariate decoding analyses to reveal early differences in the visual processing of upright and inverted letters. Compared to upright letters, inverted letters produced larger univariate N170s and had higher decoding accuracy in this period, suggesting that these increased N170 responses corresponded to a higher degree of stimuli-specific information present in the ERP signal. Univariate ERP differences were also observed earlier, starting at about 108 ms, with inverted letters producing more-positive responses than upright letters. Early decoding for upright letters began at about 80 ms and showed higher decoding accuracy than inverted letters until about 100 ms, indicating tuning for upright letters very early in the ERP time-course. This early decoding difference was specific to letters in the fovea and right visual field and was only observable from signals in left hemisphere electrodes, suggesting that this early tuning may be related to the left-to-right reading direction of our participants and the left-lateralized nature of orthographic processing in human cortex.

Appendix A: Supplemental substitution error analysis of Studies 1 and 2

When a participant makes an error in a crowding task, often times the reported target was actually one of the flankers. The fact that in many designs this occurs above chance suggests that source confusion is one of the likely mechanisms of visual crowding, although is unable to fully explain crowding (e.g. see Harrison & Bex, 2017; Freeman, Chakravarthi, and Pelli, 2012). The current studies were not designed to analyze substitution errors, given that the target-flanker spacing was controlled using a QUEST staircase, and hence, the target-flanker spacing is not held constant across participants or conditions. This makes interpreting differences between conditions with regards to their substitution error rates difficult because these differences may just be due to the differences in the target-flanker spacings. However, one benefit of using adaptive procedures for analyzing substitution errors is that each participant is guaranteed to make errors for each condition, which allows for better modeling at the item level. We have provided a supplemental analysis of substitution errors here but have left it out of the main manuscript because of the difficulty inherent with the interpretation of the results.

Study 1

The substitution error models were fit separately for study 1a and study 1b, using only trials in which an error was made. Due to the use of staircasing, errors were approximately even across conditions and participants, at about 18% (corresponding to the targeted threshold of 82%). These errors were marked as either being a substitution error (1) or not (0). A logistic mixed effects regression model was fit predicting this as the response variable. The structure of the models paralleled the crowding distance models, and included the same fixed effects of

Familiarity, Eccentricity, Hemifield, and all interactions between these variables. The random effect structure was also similar, with a random intercept for participant, and random participant slopes for the effects of Familiarity, Eccentricity, Hemifield. Unlike the crowing distance models, these models were fit at the item level so random effects for stimulus was also possible (i.e. there were repeated measures for each individual letter, inverted letter, Gabor patch). Hence, these models also included random intercepts for stimulus, and originally included random stimulus slopes for each of the fixed effects. Using the same procedure as for the crowding distance models, random slopes were removed for until the model converged. The substitution error model for study 1a was left with stimulus random slopes for the effect of Familiarity and Eccentricity, and the only stimulus random effect that could be fit with the model for study 1b was random intercepts for stimulus. Model outputs are included below in Tables S1 and S2, and plots showing the substitution error rates over the 6 spatial locations for upright letters, inverted letters, and Gabor patches are shown in Figure S1.

Study 1a model output

```
Random effects:
 Groups       Name            Variance Std.Dev.
 SUBJECT      EccentricityEff 0.22101  0.4701  
 SUBJECT.1    HemifieldEff    0.12237  0.3498  
 SUBJECT.2    CondEff         0.09619  0.3101  
 SUBJECT.3    (Intercept)     0.11691  0.3419  
 TargetStim   EccentricityEff 0.05965  0.2442  
 TargetStim.1 CondEff         0.20720  0.4552  
 TargetStim.2 (Intercept)     0.05506  0.2346  
Number of obs: 6419, groups:  SUBJECT, 60; TargetStim, 16

Fixed effects:
                                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)                           0.52456    0.09751   5.380 7.47e-08 ***
CondEff                              -0.21013    0.17801  -1.180  0.23783    
HemifieldEff                          0.08320    0.07079   1.175  0.23986    
EccentricityEff                      -0.28684    0.11111  -2.582  0.00983 ** 
CondEff:HemifieldEff                  0.30572    0.10815   2.827  0.00470 ** 
CondEff:EccentricityEff               0.99047    0.18422   5.377 7.59e-08 ***
HemifieldEff:EccentricityEff         -0.12798    0.13234  -0.967  0.33352    
CondEff:HemifieldEff:EccentricityEff -0.10065    0.26383  -0.381  0.70283    
```

Table S1. The logistic mixed effect regression model output for the substitution error analysis of Study 1a, comparing upright and inverted letters. Note that estimates and standard errors are in logits.

Study 1b model output

```
Random effects:
 Groups     Name            Variance Std.Dev.
 SUBJECT    EccentricityEff 0.08069  0.2841  
 SUBJECT.1  HemifieldEff    0.10514  0.3243  
 SUBJECT.2  CondEff         0.12739  0.3569  
 SUBJECT.3  (Intercept)     0.08527  0.2920  
 TargetStim (Intercept)     0.03629  0.1905  
Number of obs: 6787, groups:  SUBJECT, 60; TargetStim, 16

Fixed effects:
                                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)                           0.30630    0.06690   4.578 4.69e-06 ***
CondEff                               0.58128    0.11928   4.873 1.10e-06 ***
HemifieldEff                          0.03607    0.06663   0.541    0.588    
EccentricityEff                      -0.13729    0.07327  -1.874    0.061 .  
CondEff:HemifieldEff                  0.05914    0.10319   0.573    0.567    
CondEff:EccentricityEff               0.08994    0.12656   0.711    0.477    
HemifieldEff:EccentricityEff         -0.17000    0.12570  -1.352    0.176    
CondEff:HemifieldEff:EccentricityEff -0.13563    0.25104  -0.540    0.589    
```

Table S2. The logistic mixed effect regression model output for the substitution error analysis of Study 1b, comparing upright letters and Gabor patches. Note that estimates and standard errors are in logits.

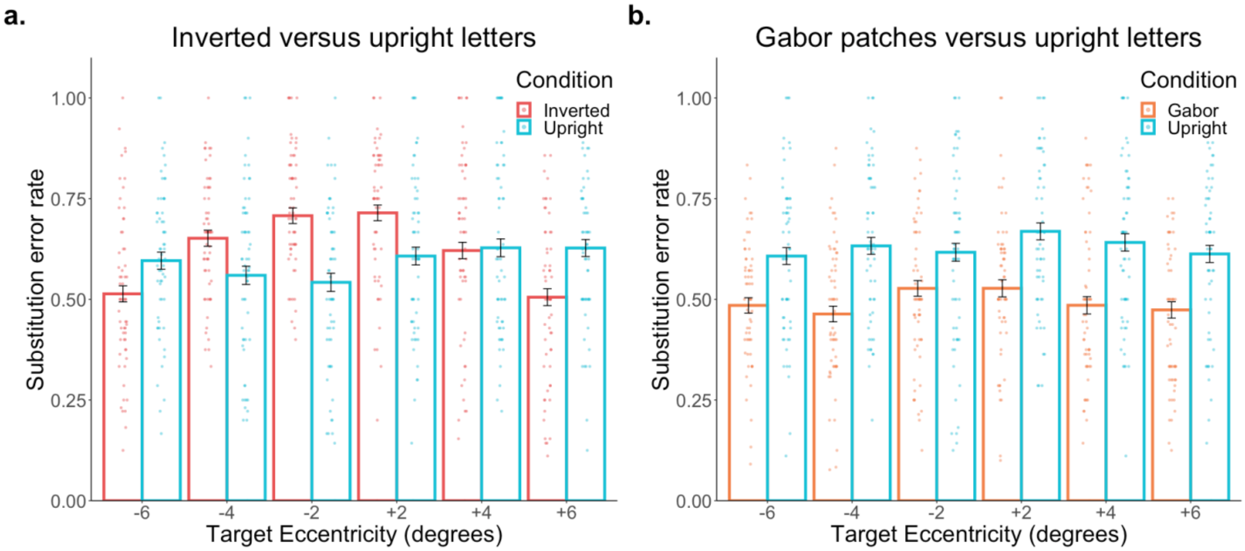

Figure S1. Average substitution error rates for each spatial location and familiarity condition, comparing upright and inverted letters in Study 1a (a) or upright letters and Gabor patches in Study 1b (b). Error bars show ±1 SEM. Dots show single-participant substitution error rates. Overall, substitution errors were made at a higher rate than chance (2/7 = ~28.5%) for all stimulus types. In study 1a, all stimuli showed higher substitution error rates at closer eccentricities (β = -0.287, SE = 0.111, p = 0.010). There was no significant difference between substitution errors made for upright and inverted letters overall (p = .238). However, inverted letters showed more eccentricity scaling than upright letters, with more substitution errors at near eccentricities and less at further eccentricities (β = 0.99, SE = 0.184, p < 0.001). Additionally, inverted letters showed more substitution errors in the left visual field while upright letters showed more substitution errors in the right visual field (β = 0.301, SE = 0.109, p = 0.005). In study 1b, upright letters had more substitution errors than Gabor patches (β = 0.581, SE = 0.119, p < 0.001). There was also a marginal effect of Eccentricity in study 1b, with more overall substitution errors at near eccentricities (β = -0.137, SE = 0.073, p = 0.061).

Study 2

The substitution error models for study 2 were similar to those of study 1. Substitution errors were predicted using a logistic mixed effects regression model. The structure of the models included fixed effects of Eccentricity, Hemifield, and the interaction between these variables. Additionally, there were fixed effects for each of the individual difference measures (spelling, author recognition, and visual working memory) and the interactions with each of these and Eccentricity and Hemifield. The random effect structure was also similar to that of Study 1, but with a few extra random effects that were able to be fit because of the large amount of data in Study 2. There were a random intercepts for participant, and random participant slopes for the effects of Eccentricity, Hemifield, and the interaction between the two. Stimulus random effects included random intercepts for stimulus, random stimulus slopes for the effects of Eccentricity, Hemifield, and the interaction between the two. The model output is included below in Table S3. A plot showing the substitution error rates over the 6 spatial locations for is shown in Figure S2, and the effects of each individual difference measure on substitution error rates is shown in Figure S3.

Study 2 model output

```
Random effects:
 Groups         Name                         Variance Std.Dev.
 SUBJECT        (Intercept)                  0.229696 0.4793
 SUBJECT.1      HemifieldEff                 0.124971 0.3535
 SUBJECT.2      EccentricityEff              0.140173 0.3744
 SUBJECT.3      HemifieldEff:EccentricityEff 0.043033 0.2074
 TargetLetter   (Intercept)                  0.008949 0.0946
 TargetLetter.1 HemifieldEff                 0.027557 0.1660
 TargetLetter.2 EccentricityEff              0.013529 0.1163
 TargetLetter.3 HemifieldEff:EccentricityEff 0.069219 0.2631
Number of obs: 17646, groups:  SUBJECT, 249; TargetLetter, 8

Fixed effects:
                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                              0.48699    0.04845  10.051  < 2e-16 ***
HemifieldEff                             0.16841    0.07158   2.353 0.018640 *
EccentricityEff                         -0.12233    0.06313  -1.938 0.052663 .
Spelling_z                              -0.12741    0.03724  -3.421 0.000624 ***
VWM_z                                   -0.13596    0.03472  -3.916    9e-05 ***
ART_z                                   -0.05413    0.03667  -1.476 0.139880
HemifieldEff:EccentricityEff            -0.20483    0.12526  -1.635 0.102012
HemifieldEff:Spelling_z                 -0.07157    0.04314  -1.659 0.097143 .
EccentricityEff:Spelling_z               0.10870    0.05056   2.150 0.031558 *
HemifieldEff:VWM_z                       0.02191    0.03998   0.548 0.583632
EccentricityEff:VWM_z                    0.06388    0.04680   1.365 0.172207
HemifieldEff:ART_z                       0.03736    0.04219   0.885 0.375915
EccentricityEff:ART_z                   -0.03647    0.05007  -0.728 0.466418
HemifieldEff:EccentricityEff:Spelling_z -0.02176    0.08819  -0.247 0.805117
HemifieldEff:EccentricityEff:VWM_z       0.16119    0.08137   1.981 0.047613 *
HemifieldEff:EccentricityEff:ART_z       0.13878    0.08783   1.580 0.114076
```

Table S3. The logistic mixed effect regression model output for the substitution error analysis of Study 2. Note that estimates and standard errors are in logits.

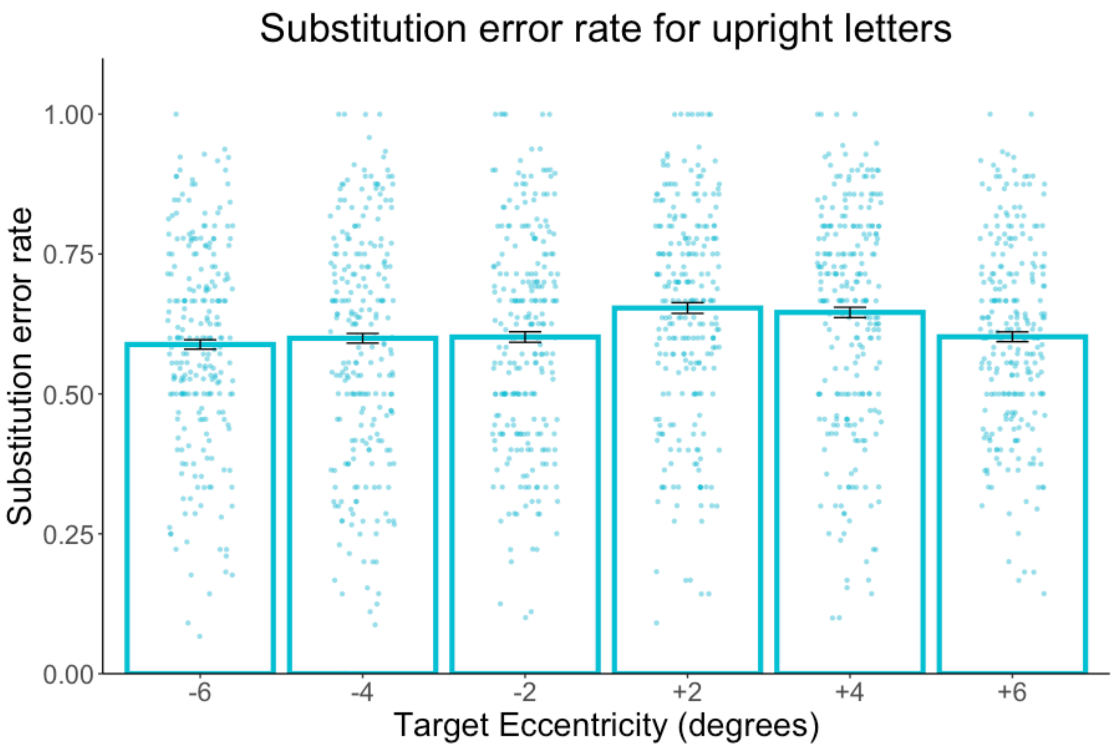

Figure S2. Average substitution error rates for each spatial location in Study 2. Error bars show ±1 SEM. Dots show single-participant substitution error rates.

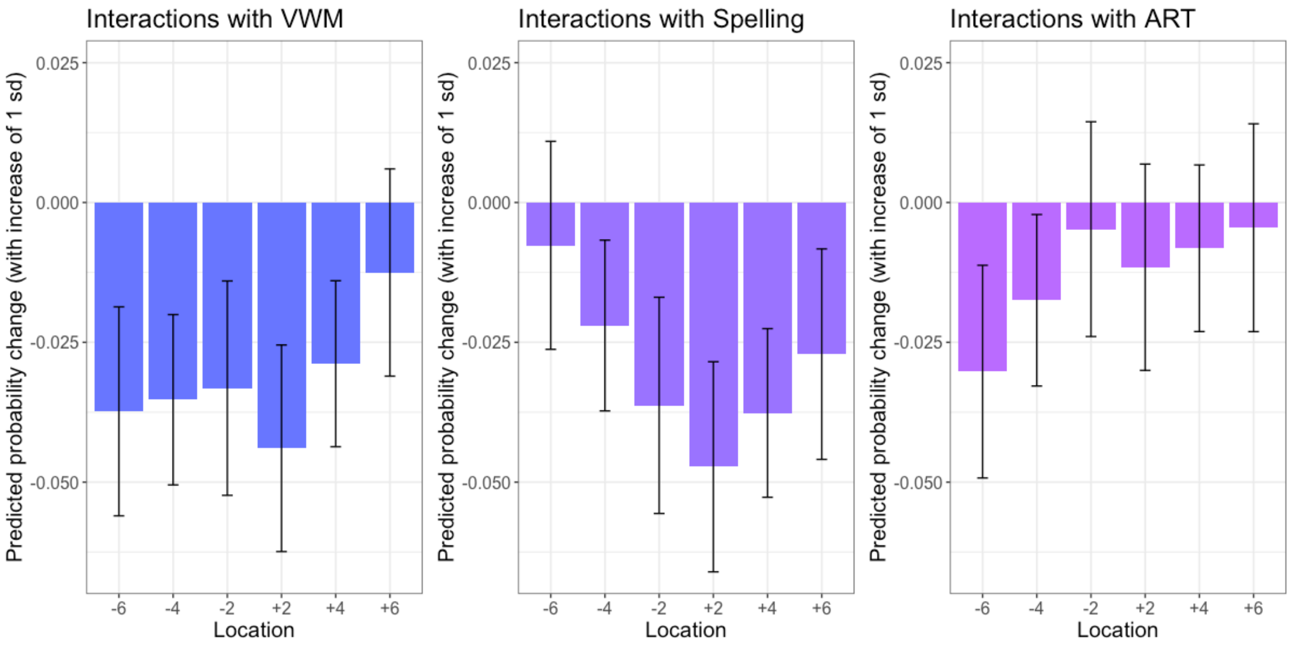

Figure S3. Model-predicted change in substitution error likelihood for an increase of 1 standard deviation on each of the individual difference measures, per location. Error bars show ±1 SEM.

As in Study 1, substitution error rates were higher than chance for upright letters in Study 2. Substitutions were more likely in the right visual field (β = 0.168, SE = 0.072, p = 0.019) and marginally significantly more likely in the near eccentricities (β = -0.122, SE = 0.063, p = 0.053). Increased scores on the spelling task predicted fewer substitution errors across all locations (β = -0.127, SE = 0.037, p < 0.001) and particularly at near eccentricities (β = 0.109, SE = 0.051, p = 0.032). Increased visual working memory predicted fewer substitution errors across all locations (β = -0.136, SE = 0.035, p < 0.001) and showed more of an eccentricity effect in the right visual field compared to the left (β = 0.161, SE = 0.081, p = 0.048).

Discussion of substitution rate analyses

All conditions in both studies showed substitution error rates above chance level (which would be 2 out of 7, or about 28.5%). This is unsurprising, since high rates of substitution errors are commonly reported, especially for stimuli like letters (e.g. Freeman, Chakravarthi, and Pelli, 2012). Fitting with this, in Study 1b we observed a lower rate of substitutions for Gabor patches compared to letters, potentially because the features differentiating Gabor patches in our design (orientation and spatial frequency) are more readily averaged or pooled. Interestingly in Study 1a, inverted letters showed eccentricity scaling for substitution errors, with higher rates closer to fixation. This could be due to the smaller crowding distances for near eccentricities, but we did not observe this pattern for upright letters, which showed the highest level of eccentricity scaling for crowding distances. This could reveal an interesting difference between upright and inverted letters with regards to substitution errors, perhaps due to interplay between experience for whole letters and letter features that are shared between upright and inverted letters. As mentioned above however, the interpretation of the substitution rates is complicated because the design was not optimized to answer these questions.

Study 2 showed high rates of substitution errors for letters as well and they were biased towards the right-near visual field similar to the overall crowding distance results. Although there was some evidence of a right hemifield bias for upright letters study 1, this was a difference in the observed effects for substitution errors for letters across the two studies. This may be due any of the differences across the two studies; namely that Study 2 used a different set of letters, a different way of responding (with the keyboard rather than the mouse), or much higher power in Study 2 (20 additional trials per condition, and 190 additional subjects). Interestingly, and also paralleling the crowding distance results, increased visual working memory and spelling ability both reduced substitution error rates. That is, these individual difference measures were associated with both smaller crowding distances and reduced substitution error rates. For visual working memory, this makes sense, since observers with greater visual working memory capacity should be able to more-precisely maintain positional information from crowded arrays, leading to reduced substitutions. For spelling, this effect was greatest at the spatial locations most important to reading, similar to the locations that spelling ability most reduced crowding distance. Since spelling ability is thought to measure precision of lexical forms, one tentative hypothesis is that this leads to more precise positional coding for letter representations, leading to a greater ability to resolve the positional source confusion that results in substitution errors. Hence, this may be one of the mechanisms by which increased spelling ability reduces crowding distance for letters. Overall, the substitution error analyses, particularly in Study 2, do provide some additional information about possible mechanisms for the ways in which reading experience, and individual differences in general, affect visual crowding. However, more research designed to test differences in substitution errors for letters and other stimuli classes as a function of observer characteristics would be very informative.

Appendix B: Supplementary model outputs for Study 2

Model reported in manuscript (n = 250 native English readers)

```
Random effects:
 Groups    Name            Variance Std.Dev.
 SUBJECT   (Intercept)     0.010516 0.10255 
 SUBJECT.1 EccentricityEff 0.003773 0.06143 
 SUBJECT.2 HemifieldEff    0.009325 0.09657 
 Residual                  0.008072 0.08984 
Number of obs: 1491, groups:  SUBJECT, 250

Fixed effects:
                                          Estimate Std. Error        df t value Pr(>|t|)    
(Intercept)                              4.343e-01  6.892e-03  2.452e+02  63.005  < 2e-16 ***
HemifieldEff                            -9.925e-02  7.685e-03  2.454e+02 -12.915  < 2e-16 ***
EccentricityEff                          2.105e-01  6.914e-03  2.444e+02  30.449  < 2e-16 ***
Spelling_z                              -2.188e-02  7.341e-03  2.458e+02  -2.980  0.00317 ** 
VWM_z                                   -3.432e-02  6.933e-03  2.458e+02  -4.950 1.38e-06 ***
ART_z                                    8.718e-04  7.237e-03  2.475e+02   0.120  0.90422    
HemifieldEff:EccentricityEff            -1.100e-02  1.144e-02  7.431e+02  -0.962  0.33630    
HemifieldEff:Spelling_z                 -1.819e-02  8.198e-03  2.467e+02  -2.218  0.02743 *  
EccentricityEff:Spelling_z               1.029e-02  7.363e-03  2.446e+02   1.397  0.16359    
HemifieldEff:VWM_z                       2.023e-02  7.741e-03  2.470e+02   2.613  0.00953 ** 
EccentricityEff:VWM_z                   -1.587e-02  6.979e-03  2.473e+02  -2.274  0.02384 *  
HemifieldEff:ART_z                      -3.852e-03  8.111e-03  2.522e+02  -0.475  0.63529    
EccentricityEff:ART_z                    1.874e-03  7.387e-03  2.593e+02   0.254  0.79991    
HemifieldEff:EccentricityEff:Spelling_z -3.658e-03  1.218e-02  7.468e+02  -0.300  0.76414    
HemifieldEff:EccentricityEff:VWM_z      -4.392e-03  1.157e-02  7.502e+02  -0.380  0.70427    
HemifieldEff:EccentricityEff:ART_z       3.072e-02  1.233e-02  7.731e+02   2.492  0.01292 *  

Type III Analysis of Variance Table with Satterthwaite's method
                                        Sum Sq Mean Sq NumDF  DenDF  F value    Pr(>F)    
HemifieldEff                            1.3464  1.3464     1 245.36 166.8019 < 2.2e-16 ***
EccentricityEff                         7.4836  7.4836     1 244.35 927.1596 < 2.2e-16 ***
Spelling_z                              0.0717  0.0717     1 245.79   8.8833  0.003167 ** 
VWM_z                                   0.1977  0.1977     1 245.76  24.4984 1.382e-06 ***
ART_z                                   0.0001  0.0001     1 247.48   0.0145  0.904215    
HemifieldEff:EccentricityEff            0.0075  0.0075     1 743.12   0.9257  0.336297    
HemifieldEff:Spelling_z                 0.0397  0.0397     1 246.74   4.9217  0.027432 *  
EccentricityEff:Spelling_z              0.0158  0.0158     1 244.58   1.9524  0.163595    
HemifieldEff:VWM_z                      0.0551  0.0551     1 246.96   6.8271  0.009529 ** 
EccentricityEff:VWM_z                   0.0417  0.0417     1 247.29   5.1698  0.023841 *  
HemifieldEff:ART_z                      0.0018  0.0018     1 252.17   0.2255  0.635286    
EccentricityEff:ART_z                   0.0005  0.0005     1 259.31   0.0644  0.799908    
HemifieldEff:EccentricityEff:Spelling_z 0.0007  0.0007     1 746.77   0.0901  0.764135    
HemifieldEff:EccentricityEff:VWM_z      0.0012  0.0012     1 750.25   0.1442  0.704274    
HemifieldEff:EccentricityEff:ART_z      0.0501  0.0501     1 773.09   6.2082  0.012925 *  
```

Table S4. The results reported in the manuscript. This model used a sample of 250 native English readers. The model had fixed effects for Hemifield (left and right coded as -0.5 and 0.5), Eccentricity (2°, 4°, and 6° degrees coded as -0.5, 0, and 0.5), Spelling (z-scored), Author Recognition (z-scored), and Visual Working Memory (z-scored). The random effect structure included random intercepts for subjects and by-subject random slopes for Hemifield and Eccentricity.

Model as preregistered (n = 267 native Latin script readers)

```
Random effects:
 Groups    Name            Variance Std.Dev.
 SUBJECT   (Intercept)     0.010873 0.10427 
 SUBJECT.1 EccentricityEff 0.004037 0.06354 
 SUBJECT.2 HemifieldEff    0.009820 0.09909 
 Residual                  0.007967 0.08926 
Number of obs: 1593, groups:  SUBJECT, 267

Fixed effects:
                                          Estimate Std. Error         df t value Pr(>|t|)    
(Intercept)                               0.436454   0.006764 262.237330  64.529  < 2e-16 ***
HemifieldEff                             -0.098438   0.007541 262.347981 -13.053  < 2e-16 ***
EccentricityEff                           0.204915   0.006733 261.641494  30.436  < 2e-16 ***
Spelling_z                               -0.023497   0.007276 262.817393  -3.229   0.0014 ** 
VWM_z                                    -0.036793   0.006812 262.797914  -5.401 1.48e-07 ***
ART_z                                    -0.001107   0.007168 264.557654  -0.154   0.8774    
HemifieldEff:EccentricityEff             -0.014730   0.010992 794.050597  -1.340   0.1806    
HemifieldEff:Spelling_z                  -0.020247   0.008123 263.704899  -2.492   0.0133 *  
EccentricityEff:Spelling_z                0.011586   0.007241 261.860387   1.600   0.1108    
HemifieldEff:VWM_z                        0.017563   0.007605 263.938367   2.310   0.0217 *  
EccentricityEff:VWM_z                    -0.014956   0.006803 264.567892  -2.199   0.0288 *  
HemifieldEff:ART_z                       -0.001149   0.008032 269.273728  -0.143   0.8863    
EccentricityEff:ART_z                     0.004983   0.007254 276.997546   0.687   0.4927    
HemifieldEff:EccentricityEff:Spelling_z  -0.001920   0.011826 797.718958  -0.162   0.8711    
HemifieldEff:EccentricityEff:VWM_z       -0.005537   0.011127 801.240322  -0.498   0.6189    
HemifieldEff:EccentricityEff:ART_z        0.030337   0.011954 825.094429   2.538   0.0113 *  

Type III Analysis of Variance Table with Satterthwaite's method
                                        Sum Sq Mean Sq NumDF  DenDF  F value    Pr(>F)    
HemifieldEff                            1.3575  1.3575     1 262.35 170.3806 < 2.2e-16 ***
EccentricityEff                         7.3803  7.3803     1 261.64 926.3360 < 2.2e-16 ***
Spelling_z                              0.0831  0.0831     1 262.82  10.4295  0.001398 ** 
VWM_z                                   0.2324  0.2324     1 262.80  29.1735 1.482e-07 ***
ART_z                                   0.0002  0.0002     1 264.56   0.0238  0.877420    
HemifieldEff:EccentricityEff            0.0143  0.0143     1 794.05   1.7956  0.180632    
HemifieldEff:Spelling_z                 0.0495  0.0495     1 263.70   6.2124  0.013300 *  
EccentricityEff:Spelling_z              0.0204  0.0204     1 261.86   2.5600  0.110807    
HemifieldEff:VWM_z                      0.0425  0.0425     1 263.94   5.3339  0.021687 *  
EccentricityEff:VWM_z                   0.0385  0.0385     1 264.57   4.8338  0.028773 *  
HemifieldEff:ART_z                      0.0002  0.0002     1 269.27   0.0205  0.886344    
EccentricityEff:ART_z                   0.0038  0.0038     1 277.00   0.4718  0.492737    
HemifieldEff:EccentricityEff:Spelling_z 0.0002  0.0002     1 797.72   0.0263  0.871096    
HemifieldEff:EccentricityEff:VWM_z      0.0020  0.0020     1 801.24   0.2476  0.618896    
HemifieldEff:EccentricityEff:ART_z      0.0513  0.0513     1 825.09   6.4407  0.011336 *  
```

Table S5. The model output if using the subject inclusion criteria specified in the preregistration. That is, including native readers of a Latin script rather than specifically native English readers. The structure of the model is identical to the one reported in the manuscript (see Supplemental Table 1), the only difference is the inclusion of 17 additional subjects. Some estimates vary slightly between the two models, but the pattern of significance is identical.

Bibliography

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. Behavior research methods, 40(1), 278-289.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. Journal of Experimental Psychology: General, 139(2), 299.

Andrews, S., Veldre, A., & Clarke, I. E. (2020). Measuring lexical quality: The role of spelling ability. Behavior Research Methods, 52, 2257-2282.

Bacigalupo, F., & Luck, S. J. (2015). The allocation of attention and working memory in visual crowding. Journal of cognitive neuroscience, 27(6), 1180-1193.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1), 289-300.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. Journal of cognitive neuroscience, 8(6), 551-565.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. Journal of cognitive neuroscience, 11(3), 235-260

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. Neuroimage, 33(2), 739-748.

Blackburne, L. K., Eddy, M. D., Kalra, P., Yee, D., Sinha, P., & Gabrieli, J. D. (2014). Neural correlates of letter reversal in children and adults. PLoS One, 9(5), e98386.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. Nature, 226(5241), 177-178.

Castaldi, E., Lunghi, C., & Morrone, M. C. (2020). Neuroplasticity in adult human visual cortex. Neuroscience & Biobehavioral Reviews, 112, 542-552.

Carrasco, C. D., Bahle, B., Simmons, A. M., & Luck, S. J. (2024). Using multivariate pattern analysis to increase effect sizes for event-related potential analyses. Psychophysiology, e14570.

Chanceaux, M., & Grainger, J. (2012). Serial position effects in the identification of letters, digits, symbols, and shapes in peripheral vision. Acta psychologica, 141(2), 149-158.

Chang, C., & Lin, C. (2011) LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27

Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early visual word processing is flexible: Evidence from spatiotemporal brain dynamics. Journal of Cognitive Neuroscience, 27(9), 1738-1751.

Chopin, A., Bediou, B., & Bavelier, D. (2019). Altering perception: the case of action video gaming. Current Opinion in Psychology, 29, 168-173.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. Neuroimage, 22(1), 466-476.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384-398.

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. Nature Reviews Neuroscience, 16(4), 234-244.

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J. B., Le Bihan, D., & Cohen, L. (2004). Letter binding and invariant recognition of masked words: behavioral and neuroimaging evidence. Psychological science, 15(5), 307-313.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., ... & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. Science, 330(6009), 1359-1364.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods, 134(1), 9-21.

Freeman, J., Chakravarthi, R., & Pelli, D. G. (2012). Substitution and pooling in crowding. Attention, Perception, & Psychophysics, 74, 379-396.

Frömer, R., Dimigen, O., Niefind, F., Krause, N., Kliegl, R., & Sommer, W. (2015). Are individual differences in reading speed related to extrafoveal visual acuity and crowding?. PloS one, 10(3), e0121986.

Grainger, J. (2018). Orthographic processing: A 'mid-level'vision of reading: The 44th Sir Frederic Bartlett Lecture. Quarterly Journal of Experimental Psychology, 71(2), 335-359.

Grainger, J., Tydgat, I., & Isselé, J. (2010). Crowding affects letters and symbols differently. Journal of Experimental Psychology: Human Perception and Performance, 36(3), 673.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology, 48(12), 1711-1725.

Harrison, W. J., & Bex, P. J. (2017). Visual crowding is a combination of an increase of positional uncertainty, source confusion, and featural averaging. Scientific reports, 7(1), 45551.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. Neuroimage, 30(4), 1383-1400.

He, Y., Legge, G. E., Yu, D.. (2013). Sensory and cognitive influences on the training-related improvement of reading speed in peripheral vision. Journal of Vision, 13 7: 14, 1– 14

He, Y., & Legge, G. E. (2017). Linking crowding, visual span, and reading. Journal of vision, 17(11), 11-11.

Hussain, Z., Webb, B. S., Astle, A. T., & McGraw, P. V. (2012). Perceptual learning reduces crowding in amblyopia and in the normal periphery. Journal of Neuroscience, 32(2), 474-480.

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., Luck, S. J., & Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. Neuropsychology, 27(2), 220–229. https://doi.org/10.1037/a0032060

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. Journal of neurophysiology, 58(6), 1233-1258.

Joo, S. J., White, A. L., Strodtman, D. J., & Yeatman, J. D. (2018). Optimizing text for an individual's visual system: The contribution of visual crowding to reading difficulties. Cortex, 103, 291-301.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. Philosophical Transactions of the Royal Society B: Biological Sciences, 361(1476), 2109-2128.

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” Journal of Statistical Software, 82(13), 1–26.

Kurzawski, J. W., Burchell, A., Thapa, D., Winawer, J., Majaj, N. J., & Pelli, D. G. (2021). The Bouma law accounts for crowding in fifty observers. BioRxiv, 2021-04.

Kwon, M., Legge, G. E., & Dubbels, B. R. (2007). Developmental changes in the visual span for reading. Vision research, 47(22), 2889-2900.

Kyllingsbaek, S., & Bundesen, C. (2009). Changing change detection: Improving the reliability of measures of visual short-term memory capacity. Psychonomic Bulletin & Review, 16, 1000–1010. https://doi.org/10.3758/PBR.16.6.1000

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. Vision research, 48(5), 635-654.

Li, W. (2016). Perceptual learning: Use-dependent cortical plasticity. Annual Review of Vision Science, 2, 109-130.

Ling, S., Lee, A. C., Armstrong, B. C., & Nestor, A. (2019). How are visual words represented? Insights from EEG-based visual word decoding, feature derivation and image reconstruction. Human brain mapping, 40(17), 5056-5068.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. Frontiers in human neuroscience, 8, 213.

Luck, S. J. (2014). An introduction to the event-related potential technique. MIT press.

Lui, K. F., Lo, J. C., Maurer, U., Ho, C. S. H., & McBride, C. (2021). Electroencephalography decoding of Chinese characters in primary school children and its prediction for word reading performance and development. Developmental Science, 24(3), e13060.

Malania, M., Pawellek, M., Plank, T., & Greenlee, M. W. (2020). Training-induced changes in radial–tangential anisotropy of visual crowding. Translational Vision Science & Technology, 9(9), 25-25.

Manassi, M., & Whitney, D. (2018). Multi-level crowding and the paradox of object recognition in clutter. Current Biology, 28(3), R127-R133.

Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., ... & Brandeis, D. (2006). Coarse neural tuning for print peaks when children learn to read. Neuroimage, 33(2), 749-758.

Maurer, U., & McCandliss, B. D. (2007). The development of visual expertise for words: the contribution of electrophysiology. In Single-word reading (pp. 57-77). Psychology Press.

McCarron, S. P., & Kuperman, V. (2021). Is the author recognition test a useful metric for native and non-native English speakers? An item response theory analysis. Behavior Research Methods, 53(5), 2226-2237.

Michael, G. A., & Ojéda, N. (2005). Visual field asymmetries in selective attention: Evidence from a modified search paradigm. Neuroscience Letters, 388(2), 65-70.

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. Behavior research methods, 47, 1095-1109.

Nara, S., Raza, H., Carreiras, M., & Molinaro, N. (2023). Decoding numeracy and literacy in the human brain: insights from MEG and MVPA. Scientific Reports, 13(1), 10979.

Pegado, F., Comerlato, E., Ventura, F., Jobert, A., Nakamura, K., Buiatti, M., ... & Dehaene, S. (2014). Timing the impact of literacy on visual processing. Proceedings of the National Academy of Sciences, 111(49), E5233-E5242.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. Journal of neuroscience methods, 162(1-2), 8-13.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. Vision research, 46(28), 4646-4674.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. Nature neuroscience, 11(10), 1129-1135.

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. Journal of vision, 7(2), 20-20.

Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. Trends in cognitive sciences, 15(6), 246-253.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. The quarterly journal of experimental psychology, 62(8), 1457-1506.

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. Psychonomic bulletin & review, 17(6), 834-839.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. Journal of Experimental Psychology. Human Perception and Performance, 37, 514–528.

Ramamurthy, M., White, A. L., Chou, C., & Yeatman, J. D. (2021). Spatial attention in encoding letter combinations. Scientific Reports, 11(1), 1-12.

Risse, S. (2014). Effects of visual span on reading speed and parafoveal processing in eye movements during sentence reading. Journal of Vision, 14, 1–13.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport, 11(1), 69-72.

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. Neuroimage, 20(3), 1609-1624.

Sagi, D. (2011). Perceptual learning in Vision Research. Vision Research, 51(13), 1552–1566. https://doi.org/10.1016/j.visres.2010.10.019

Schotter, E.R., Angele, B. & Rayner, K. Parafoveal processing in reading. Atten Percept Psychophys 74, 5–35 (2012). https://doi.org/10.3758/s13414-011-0219-2

Schroeder, C. E., Mehta, A. D., & Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cerebral cortex, 8(7), 575-592.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. Reading research quarterly, 402-433

Stevens, C., McIlraith, A., Rusk, N., Niermeyer, M., & Waller, H. (2013). Relative laterality of the N170 to single letter stimuli is predicted by a concurrent neural index of implicit processing of letternames. Neuropsychologia, 51(4), 667-674.

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., & Cohen, L. (2011). Specialization for written words over objects in the visual cortex. Neuroimage, 56(1), 330-344.

Szwed, M., Qiao, E., Jobert, A., Dehaene, S., & Cohen, L. (2014). Effects of literacy in early visual and occipitotemporal areas of Chinese and French readers. Journal of cognitive neuroscience, 26(3), 459-475.

Thesen, T., McDonald, C. R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., ... & Halgren, E. (2012). Sequential then interactive processing of letters and words in the left fusiform gyrus. Nature communications, 3(1), 1284.

Toussi, N., Takai, O., Bann, S., Rowe, J., Hildebrand, A. J., & Herdman, A. (2023). Combined studies of N170/M170 responses to single letters and pseudoletters. bioRxiv, 2023-10.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. Vision research, 42(20), 2357-2369.

Tydgat, I., & Grainger, J. (2009). Serial position effects in the identification of letters, digits, and symbols. Journal of Experimental Psychology: Human Perception and Performance, 35(2), 480.

Uno, T., Kasai, T., & Seki, A. (2024). The Developmental Change of Print-Tuned N170 in Highly Transparent Writing Systems 1. Japanese Psychological Research, 66(1), 82-89.

US Office of Management and Budget. (2016, September 30). Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity. Federal Register. https://www.federalregister.gov/documents/2016/09/30/2016-23672/standards-for-maintaining-collecting-and-presenting-federal-data-on-race-and-ethnicity

van Zanden, J., Baten, J., Mira d'Ercole, M., Rijpma, A., & Timmer, M. P. (2014). How was life?: Global well-being since 1820. OECD Publishing. doi: 10.1787/9789264214262.

Vejnović, D., & Zdravković, S. (2015). Side flankers produce less crowding, but only for letters. Cognition, 143, 217-227.

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. Quarterly Journal of Experimental Psychology, 67(4), 703-727.

Veldre, A., & Andrews, S. (2016). Semantic preview benefit in English: Individual differences in the extraction and use of parafoveal semantic information. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(6), 837.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. Perception & psychophysics, 33(2), 113-120.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. Trends in cognitive sciences, 15(4), 160-168.

Winsler, K., Grainger, J., & Holcomb, P. J. (2022). On letter-specific crowding and reading: Evidence from ERPs. Neuropsychologia, 176, 108396.

Winsler, K., Holcomb, P. J., & Emmorey, K. (2023). Electrophysiological patterns of visual word recognition in deaf and hearing readers: an ERP mega-study. Language, cognition and neuroscience, 38(5), 636-650.

Yeatman, J. D., & White, A. L. (2021). Reading: the confluence of vision and language. Annual Review of Vision Science, 7, 487-517.

Zhang, G., Carrasco, C. D., Winsler, K., Bahle, B., Cong, F., & Luck, S. J. (2024). Assessing the effectiveness of spatial PCA on SVM-based decoding of EEG data. NeuroImage, 293, 120625.

Zhang, W., & Kappenman, E. S. (2024). Maximizing signal-to-noise ratio and statistical power in ERP measurement: Single sites versus multi-site average clusters. Psychophysiology, 61(2), e14440.

Zhao, C., Vogel, E., & Awh, E. (2022). Change localization: A highly reliable and sensitive measure of capacity in visual working memory. Attention, Perception, & Psychophysics. https://doi.org/10.3758/s13414-022-02586-0