Visual word recognition relies on an orthographic prediction error signal

Benjamin Gagl[1,2*], Jona Sassenhagen[1], Sophia Haan[1], Klara Gregorova[1], Fabio Richlan[3] and Christian J. Fiebach[1,2,4]

[1] Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt/Main, Germany

[2] Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt/Main, Germany

[3] Centre for Cognitive Neuroscience, University of Salzburg, Hellbrunnerstrasse 34, 5020 Salzburg, Austria.

[4] Brain Imaging Center, Goethe University Frankfurt, Frankfurt/Main, Germany

* Corresponding author

Abstract

Current cognitive models of reading assume that word recognition involves the 'bottom-up' assembly of perceived low-level visual features into letters, letter combinations, and words. This rather inefficient strategy, however, is incompatible with neurophysiological theories of Bayesian-like predictive neural computations during perception. Here we propose that prior knowledge of the words in a language is used to 'explain away' redundant and highly expected parts of the percept. As a result, subsequent processing stages operate upon an optimized representation highlighting information relevant for word identification, i.e., the orthographic prediction error. We demonstrate empirically that the orthographic prediction error accounts for word recognition behavior. We then report neurophysiological data showing that this informationally optimized orthographic prediction error is processed around 200 ms after word-onset in the occipital cortex. The remarkable efficiency of reading, thus, is achieved by optimizing the mental representation of the visual percept, based on prior visual-orthographic knowledge.

Introduction

Written language – script – developed over the last ~8,000 years in many different variants (Haarmann, 2007). It is a symbolic representation of meaning, based on the combination of simple high contrast visual features (oriented lines) into ultimately linguistically meaningful units. All current cognitive-psychological reading models (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007) assume that the recognition of written words involves the 'bottom-up' assembly of such visual line representations into abstract orthographic representations of letters, letter combinations (e.g., bi- or trigrams), and ultimately words. The most widely-accepted brain-based account of reading (Dehaene & Cohen, 2011; Dehaene, Cohen, Sigman, & Vinckier, 2005) proposes that this is realized by perceptual neurons that – starting from line-orientation sensitive neurons in primary visual cortex – represent successively more complex combinations of lines. Higher-order visual neurons in the ventrotemporal part of the left brain hemisphere are thought to provide the 'access code' for retrieving word meaning during comprehension (Coltheart et al., 2001; Perry et al., 2007).

In steep contrast, a currently prevalent model questions the reliance of visual information processing on a purely bottom-up signal (e.g. Clark, 2013; K. Friston, 2005; Rao & Ballard, 1999; Srinivasan, Laughlin, & Dubs, 1982), and favors a top-down guided, active and prediction-based approach to visual perception. From the point of view of this predictive coding perspective, exclusively bottom-up, feature-based sensory processing is inefficient – and thus hard to integrate with the empirical finding that most of us read at a remarkable speed (Gagl et al., 2018; Kliegl, Nuthmann, & Engbert, 2006; Rayner, 2009). Nevertheless, there is so far no plausible account for the remarkable efficiency of sensory processing of visual words, as a basis for fast visual word recognition.

Here we adopt the domain-general neurophysiological theory of predictive coding to visual word recognition and reading. Predictive coding postulates that perceived regularities in the world are used to build up internal models of the (hidden) causes of sensory events, and that these internal predictions are imprinted in a top-down manner upon the hierarchically lower sensory systems, making sensory analysis of perfectly expected inputs, in the best case, obsolete (K. Friston, 2005; Rao & Ballard, 1999). In recent years, this framework has received support in many domains of perceptual neuroscience from retinal coding (Srinivasan et al., 1982), auditory perception (Todorovic, van Ede, Maris, & de Lange, 2011; Wacongne, Changeux, & Dehaene, 2012) and speech perception (Arnal, Wyart, & Giraud, 2011;

Gagnepain, Henson, & Davis, 2012) to object (Kersten, Mamassian, & Yuille, 2004) and face recognition (Schwiedrzik & Freiwald, 2017). According to predictive coding theory, sensory percepts that confirm contextual or knowledge-based expectations elicit relatively reduced neuronal responses. Predictive processing, thus, increases the resource-efficiency of perception. In contrast, when new input violates these expectations, a prediction error signal is generated (e.g. Todorovic et al., 2011) which, according to current theorizing, signals the unexpected part of the percept to higher levels of cortical processing, where it is used for model updating and thus optimizing future predictions (Clark, 2013; Rao & Ballard, 1999).

If predictive coding is indeed a general principle of brain function, it should apply also to the perceptual processes involved in visual word recognition. Previously, (Price & Devlin, 2011) proposed a role for predictive processes during higher stages of word recognition, particularly related to the interactive integration of contextual linguistic information (e.g., semantic or phonological) with visual-spatial bottom-up information, in the service of word identification. In contrast, we here focus on earlier stages of sensory processing of visual words. Remember that one of the main claims of predictive coding is that our brain 'explains away' redundant (and thus non-informative) aspects of the sensory percept, to optimize information processing at the lowest levels of perception. We here propose that during visual word recognition, this mechanism operates on the basis of the orthographic knowledge that we have acquired about language. Interestingly, the feature-configurations that constitute letters and words, i.e., that are part of our orthographic knowledge of language, contain highly redundant information (Changizi, Zhang, Ye, & Shimojo, 2006) – like vertical lines often occurring at the same position (e.g., the left vertical line in E, R, N, P, B, D, F, H, K, L, M) or letters often positioned at the same location in a word (e.g., *s* or *y* as final letters in English). As redundancies contribute very little to unique letter and word identification, we here propose that using prior orthographic knowledge to predict away the redundant part of the percept is a neurophysiologically plausible strategy of our brain to reduce the amount of to-be-processed information – and thus a plausible way of increasing the efficiency of visual word recognition, relative to strict feature-based bottom-up processing of the full visual input. At the same time, we assume that the subsequent abstractions of letter and word representations, as assumed by current visual word recognition theories (e.g. Coltheart et al., 2001; Dehaene et al., 2005), can operate upon an informationally optimized representation of the visual input. The proposed orthographic prediction model of reading, thus, is not in principle incompatible with established models of visual word processing and reading.

To test which sensory processing hypothesis is adequate for visual word recognition behavior and related neuronal activation, we here report an implementation of the proposed prediction-based word recognition model, and use it to compare two parameters, one reflecting strictly bottom-up visual processing and one based on a top down-/prediction-based optimization of the sensory representation of the perceived stimulus word.

A Predictive Coding Model of Visual Input Optimization in Reading

We postulate that the brain identifies words not on the basis of the full (i.e., 'bottom-up') physical input into the visual system contained in a string of letters, but rather based on an informationally optimized representation of the percept that only reflects the non-predictable, i.e., informative part of the input (Fig. 1a). In the predictive coding framework, this non-redundant portion of a stimulus is formalized as a prediction error; we thus propose that during visual word recognition, internal (i.e., knowledge- or context-dependent) visual-orthographic predictions are used to informationally optimize the sensory input, so that further processing stages operate upon an *orthographic prediction error* (oPE) signal. It is commonly believed that higher level linguistic representations can initiate specific expectations about upcoming words (DeLong, Urbach, & Kutas, 2005; Kliegl et al., 2006; Nieuwland et al., 2018; Price & Devlin, 2011) – e.g., about the class (noun or verb) and meaning of the next word in a sentence like *"The scientists made an unexpected … (discovery)"*. The fundamental difference between these psycholinguistic assumptions about semantic and syntactic predictions and the proposed *visual-orthographic prediction (VOP) model* of reading, postulates predictive processes already at much earlier stages of visual processing. While this is in line with general considerations about information optimization as fundamental property of perceptual systems (Niven & Laughlin, 2008), this proposal differs radically from current (neuro-)cognitive models of reading – which rely on the full bottom-up visual-sensory input.

Here, we demonstrate a quantitative implementation of this model for the most frequently investigated paradigm in reading research, single word recognition. In the absence of sentence context, the redundant visual information (i.e., the *visual-orthographic prediction* or, in Bayesian terms, the *prior*) is a function of our orthographic knowledge of words. We here approximate this knowledge quantitatively as the pixel-by-pixel mean over image representations of all words derived from a psycholinguistic database (Brysbaert et al., 2011;

see Fig. 1b and Methods). Interestingly, the resulting visual-orthographic predictions look similar across different languages sharing the same writing system (compare Fig. 1b,c) and, when compared directly, correlations of gray values are high ranging from .95 to .99.

To empirically examine the assumptions of the VOP model, we estimate the orthographic prediction error as a pixel-by-pixel subtraction of this visual-orthographic prediction from each perceived word (Fig. 1d). This step of 'predicting away' the redundant part of written words reduces the amount of to-be-processed information by up to 51% (on average 33%, 37%, and 31% for English, French, and German, respectively; see Methods, Formula 4), thereby optimizing the visual input signal in the sense of highlighting only its informative parts (Fig. 1d). According to the VOP model, the resulting orthographic prediction error is a critical early stage of word identification, representing the access code that our brain uses to activate word meaning. In the following, we provide empirical support for this model by demonstrating that the orthographic prediction error (i) is correlated with orthographic familiarity of words measured as a property of lexicon statistics, (ii) accounts for response times in three languages, (iii) is represented in occipital brain regions, and (iv) is active at around ~170 ms after word onset.

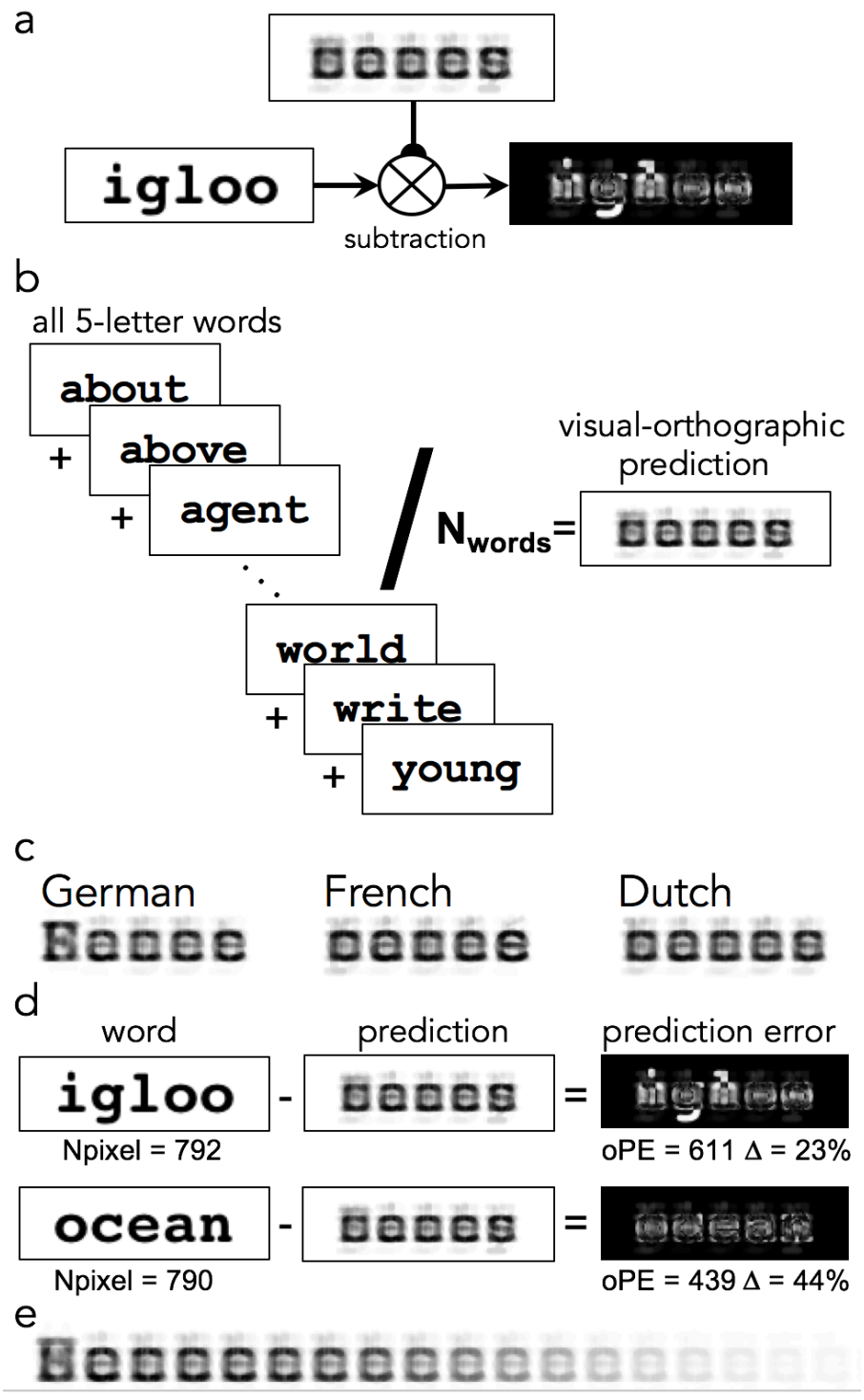


Figure 1. Visual-orthographic prediction (VOP) model of reading. (a) The VOP model assumes that during word recognition, redundant visual-orthographic information is 'explained away', thereby highlighting the informative aspects of the percept. Subsequent stages of word recognition and linguistic processing (i.e. accessing abstract letter and word representations), thus, operate upon an informationally optimized input representation. This assumption is here tested for single-word reading, i.e., independent of context, by subtracting a 'visual-orthographic prediction' from the input. (b) The visual-orthographic, knowledge-based prediction is implemented as a pixel-by-pixel mean across image

representations of all known words (here approximated by all words in a psycholinguistic database; only five letters words, as in most experiments reported here; but see Supplemental figure 1a for a prediction including different word lengths). The resulting visual-orthographic prediction, shown on the right, contains the most redundant visual information across all words. (c) Across multiple languages, these predictions are very similar, with the exception of the upper-case initial letter that is visible in the German prediction (because experiments in German involved only nouns). (d) The orthographic prediction error (oPE) is estimated, for each word, by a pixel-by-pixel subtraction of the orthographic prediction from the input word (based on their image representations; see Methods for details). While the two example words have similar numbers of pixels, subtracting the orthographic prediction results in substantially different residual (i.e., oPE) images. The values underneath the prediction error images represent a quantitative estimate of the orthographic prediction error, the sum of non-black pixels per image, and show that the amount of information reduction (Δ) can differ strongly between words. (e) Letter-length unspecific prediction for German, based on ~190.000 words.

Results

Lexicon-based Characterization of the Orthographic Prediction Error

Cognitive psychologists have developed a number of quantitative measures to characterize words (Brysbaert et al., 2011; Coltheart, Davelaar, Jonasson, & Besner, 1977; Yarkoni, Balota, & Yap, 2008), mostly derived from large text corpora and psycholinguistic word databases (Heuven, Mandera, Keuleers, & Brysbaert, 2014; Keuleers, Brysbaert, & New, 2010; see Fig. 2a for most important characteristics and examples). Abundant empirical research demonstrates that these lexicon-based word characteristics are predictive of different aspects of reading behavior (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Rayner, 2009). Accordingly, understanding how the orthographic prediction error derived from the implemented VOP model (see Fig. 1) relates to these measures provides an important first indication that this informationally optimized perceptual signal is critically involved in word recognition.

a

Orthographic measure	Example: *read*	Value
Coltheart neighbors (N)	real, head, road, lead, rear …	11
Orthographic Levenshtein distance (OLD)	real: 1, ready: 1, red: 1, great: 2, bad: 2, …; Mean distance over the 20 nearest words	1
Bigram frequency	re, ea, ad; summed occurrences in lexicon	29951
Trigram frequency	rea, ead; summed occurrences in lexicon	4528
Quadrigram frequency	read; summed occurrences in lexicon	90
Word frequency	Log. occurrences per million	5.3

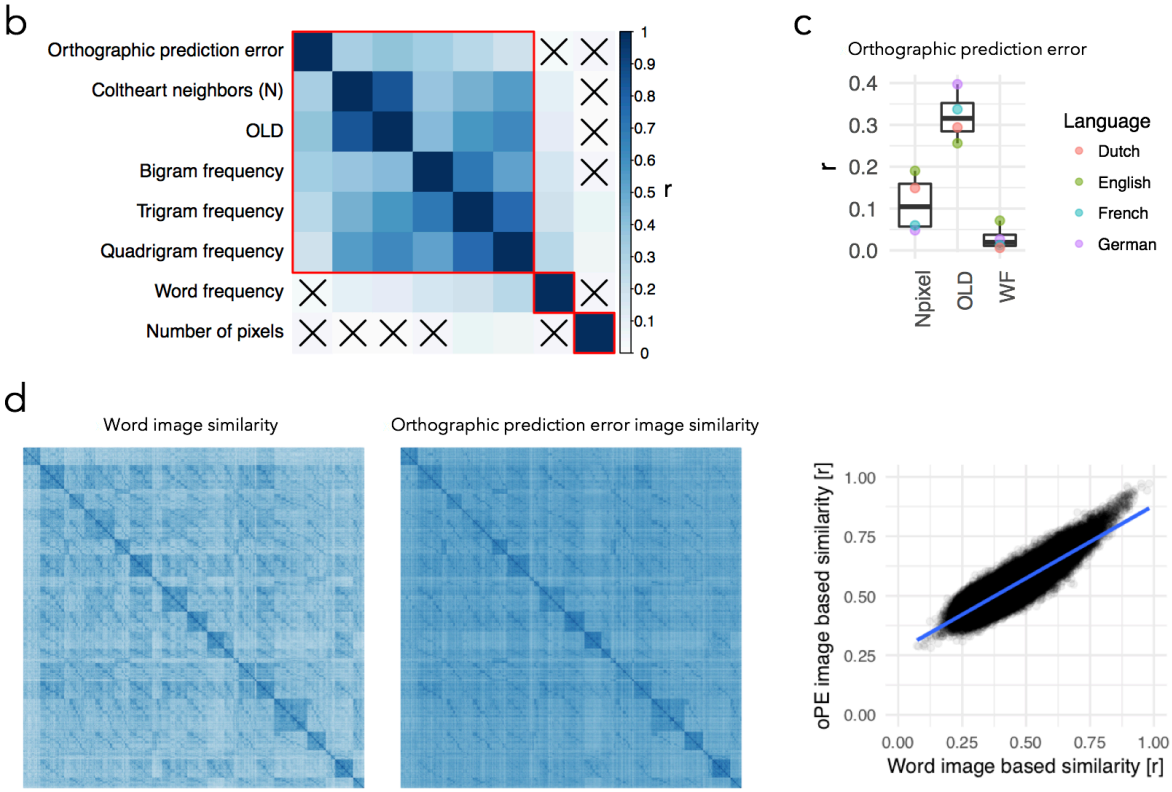


Figure 2. Comparison of orthographic prediction error to established lexicon-based word characteristics. (a) Overview of established word characteristics, exemplified for the word 'read': Coltheart's neighborhood size (Coltheart N; Coltheart et al., 1977), orthographic Levenshtein distance (OLD20; Yarkoni et al., 2008), sub-lexical frequency measures (bi-, tri-, and quadri-gram frequencies, i.e. number of occurrences of two, three, and four-letter combinations from the target word, in the lexicon), and word frequency as calculated from established linguistic corpora (see Methods for details). (b) Clustered correlation matrix between the orthographic prediction error, the number of pixels per original image, which represents an estimate of the pure amount of physical bottom-up input in the present study, and the described word characteristics (cf. panel a), applied to a set of 3,110 German nouns. Red rectangles mark clusters and black crosses mark non-significant correlations ($p < .05$; Bonferroni corrected to $p < .00179$). (c) Correlations between the orthographic prediction error and number of pixels per word (Npixel),

orthographic similarity (OLD20), and word frequency (WF), for four different languages. (d) Representational similarity matrices (RSM; Kriegeskorte, Mur, & Bandettini, 2008) for original word images (left panel) and orthographic prediction error images (central panel). Each similarity matrix reflects the correlations among the gray values of all 3,110 words (in total 9,672,100 correlations per matrix). Note, the words were sorted alphabetically. The color scale indexing correlation strength is equivalent to the one used in (b). The right panel shows the correlation between word- and orthographic prediction error-based RSMs.

Across all words, the orthographic prediction error (calculated as a summed difference between the actual stimulus and the knowledge-based visual-orthographic prediction; cf. Fig. 1d and Methods) clusters with several measures commonly interpreted as orthographic (Fig. 2b). These classic orthographic characteristics reflect the (non-) uniqueness of words in terms of their orthographic similarity to other words (e.g., the number of Coltheart neighbors Coltheart et al., 1977 or the orthographic distance OLD20; Yarkoni et al., 2008 to other words; cf. Fig. 2a) and letter co-occurrences (e.g., bi- and trigram frequency; cf. Fig. 2a). Note that these measures describe the statistics of letters and letter combinations in relation to all words retrieved from a lexicon database (Keuleers, Brysbaert, et al., 2010); in cognitive psychological research, they are consistently being associated with the first, i.e., orthographic stages of processing written words (Coltheart et al., 2001; Grainger & Jacobs, 1996). This is an important result as it demonstrates that a neurophysiologically inspired transformation of the visual stimulus that optimizes its information content, i.e., the here-proposed orthographic prediction error, is meaningfully related to orthographic properties of words as derived from lexicon-based statistics.

In contrast, the orthographic prediction error is not correlated with the frequency of occurrence of a word in a language (Fig. 2b), which is typically taken as indicator of the difficulty of the process of accessing word meaning on the basis of an already-decoded orthographic access code (Coltheart et al., 2001). This dissociation between the orthographic prediction error and word frequency replicates across languages (Fig. 2c) and is in fact much more pronounced for the orthographic prediction error than the so-far predominant measures of orthographic similarity and orthographic neighborhood (Fig. 2b). Only two classical orthographic measures (tri- and quadrigram frequency) were weakly correlated with the raw pixel count of the words (Fig. 2b), providing first evidence that the neurophysiologically inspired orthographic prediction error is more important for a mechanistic understanding of reading than the full physical input contained in a printed word, without losing the ability of

discriminating the word identities. The latter was indicated by a strong correlation (r = .87) between the representational similarity matrices of the word and orthographic prediction error images (Fig. 2d). Here, the gray values from each word/orthographic prediction error image were correlated, pixel by pixel, separately for the original words and the orthographic prediction error images. The resulting so-called representational similarity matrices reflect the similarity structures among the entire set of items (Edelman, 1998; Kriegeskorte et al., 2008). The strong correlation of the word image and prediction error image similarity matrices, which is directly visible when visually comparing the structure of the two matrices, indicates that the representational similarity structure, or in other words the discriminability between items, is preserved after perceptual optimization of the sensory input as proposed by the VOP model.

Accounting for Word Recognition Behavior

As a next empirical test of the visual-orthographic prediction of reading, we evaluated how well the orthographic prediction error performs in accounting for behavior in an established and widely-used word recognition task. 35 human participants were asked to decide as fast as possible by button press whether written letter-strings (presented on the computer screen; 1,600 items; 5 letters length; language: German) were words or not (lexical decision task). The orthographic prediction error represents the deviance of a given letter-string from our knowledge-based orthographic expectation, and thus how (un-)likely it is that the given letter-string constitutes a word. Accordingly, participants should be fast in identifying letter-strings with low orthographic prediction error as words and fast in rejecting non-words with a high orthographic prediction error.

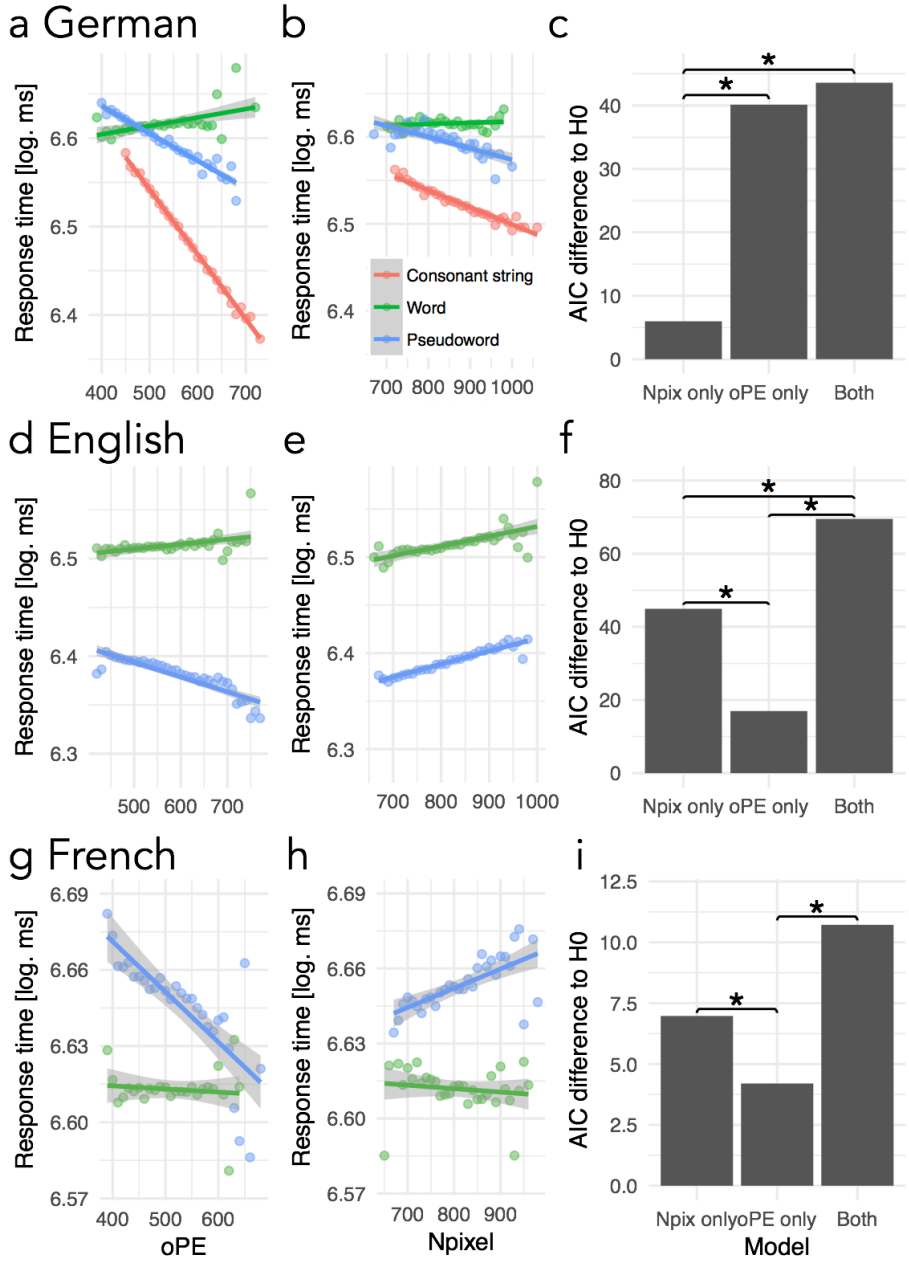


Figure 3. Word/non-word decision task behavior. (a) Orthographic prediction error (oPE) and (b) number of pixels (Npixel) effects on response times in a word/non-word decision task (German 5-letter nouns; overall error rate 7.4%; see Supplemental table 1 for all detailed statistical analysis). Green lines show the effects for words, blue lines for pseudowords (pronounceable non-words), and red lines for consonant strings (unpronounceable non-words). Dots represent mean reaction time estimates across all participants, separated into bins of oPE (width of 10) and stimulus category, after excluding confounding effects. (c) Results from model comparisons. First all models are compared to the null model with only word/non-word status and word frequency as predictors. Subsequently, a model adding only the oPE predictor, a model adding only the Npixel predictor, and one model adding both predictors to the predictors of the null model, were compared to the null model. Note that also the interaction terms with the word/non-word parameter were

included. The Akaike Information Criterion (AIC) difference to the null model is shown for the three models. A positive value represents an increase in model fit; asterisks mark significant differences ($p < .05$ Bonferroni corrected for multiple comparisons; 6 comparisons, three in relation to the null model and three comparing the alternative models; corrected significance threshold $p < .0083$). (d-f) Analogous results for English and (g-i) for French word/non-word decision tasks.

Fig. 3a shows exactly this pattern of response times, i.e., a word/non-word by orthographic prediction error interaction (linear mixed model/LMM estimate: 0.03; SE = 0.01; $t = 5.0$; see Methods for details on linear mixed effects modeling and Supplemental table 1 for detailed results). No significant interaction or fixed effect of the number of pixels estimate (i.e. the sum of all black pixels contained in a word), representing the strictly bottom-up model, was found (Fig. 3b; Interaction: estimate: 0.00; SE = 0.01; $t = 0.0$; Fixed effect: -0.01; SE = 0.00; $t = 1.8$). To directly compare if the response times are more adequately described by the VOP model, represented by the orthographic prediction error, or by the pure bottom up model, represented by the number of pixels predictor for each word, we performed an explicit model comparison (see Methods for details) of four models: the full model, including as predictors the orthographic prediction error and the number of pixels, a pure prediction error model, a pure number of pixels model, and a null model without any of the two predictors. Fig. 3c shows that, in contrast to the null model, the three alternative models showed higher model fits (all χ^2's > 9.9; all p's < .007; Bonferroni corrected p threshold: 0.0083), but this increase was significantly larger for the models including the orthographic prediction error. In addition, the model including only the orthographic prediction error explained substantially higher amounts of variance when compared to the model including only the number of pixels parameter (AIC difference: 34; $\chi^2(0) = 34.2$; $p < .001$) with no substantial increase for the combined model (AIC difference: 3; $\chi^2(2) = 7.2$; $p = .02$). This indicates that the bottom-up model explains substantially less variance in word recognition behavior than the orthographic prediction error based model.

Additionally, including orthographic distance (OLD20; Yarkoni et al., 2008) as predictor improved the model fit further (AIC difference comparing the full model with and without OLD20: 104 $\chi^2(2) = 105.8$; $p < .001$) but did not affect the significance of the word/non-word-by-orthographic prediction error interaction (Interaction effect estimate after including additional parameters: 0.03; SE = 0.01; $t = 5.2$). This indicates that despite its correlation with other orthographic measures (Figure 2b, c), the orthographic prediction error

accounts for unique variance components in word recognition behavior that cannot be explained by other word characteristics.

We also replicate this interaction when calculating the orthographic prediction error using a length-unspecific visual-orthographic prediction (i.e., based on all ~190,000 German words from the SUBTLEX database Brysbaert et al., 2011; 2-36 letters length; cf. Fig. 1e; LMM estimate of interaction effect: 0.03; SE = 0.01; t = 4.5; for a replication in English and a more extensive investigation of the interaction effect for multiple word lengths see Supplemental figure 1b). Interestingly, length-specific and length-unspecific orthographic prediction errors are highly correlated (e.g., German: r = .97), showing that the prediction-based word recognition process proposed by the VOP model is a general mechanism independent of word length constraints. This notion is in line with findings from natural reading, which show that low level visual features like length can be extracted from parafoveal vision prior to fixating the word (Cutter, Drieghe, & Liversedge, 2014; Gagl, Hawelka, Richlan, Schuster, & Hutzler, 2014; Schotter, Angele, & Rayner, 2012). The use of a fixed of word length in our German lexical decision experiment is therefore not necessarily artificial, since in natural reading word length is known prior to fixation. Combined, these results demonstrate that the orthographic prediction error is meaningfully related to word recognition behavior and independent of word length.

Generalization across languages

The interaction effect shown for German could be replicated in two open datasets from other languages (British English, 78 participants and 8,488 words/non-words: Fig. 3d; estimate: 0.008; SE = 0.002; t = 4.2, Keuleers, Lacey, Rastle, & Brysbaert, 2012; French, 974 participants and 5,368 words/non-words: Fig. 3g; estimate: 0.005; SE = 0.002; t = 2.0, Ferrand et al., 2010; see Fig. 3; see also Supplemental figure 2 for two further datasets from Dutch and Supplemental table 1 for detailed results of English and French). However, in contrast to German, in both datasets we also found a significant effect of the number of pixels parameter (Fig. 3e,h; British: fixed effect: 0.008; SE = 0.001; t = 6.7; French: interaction with word/non-word status: -0.007; SE = 0.002; t = 3.0). In terms of model comparison, the pattern derived from German, i.e., strongest model fit increases when the orthographic prediction error was included, could not be recovered for English and French. Rather, we found that the role of the number of pixels parameter for describing the response times was larger than in German (see Fig 3f,i). Still, the combined model showed the best model fit in all three languages (although this difference was not significant for French after Bonferroni correction:

$\chi^2(2) = 7.8$; $p = .02$) indicating that both the orthographic prediction error and the number of pixels parameter are relevant in explaining word/non-word decision behavior (oPE only vs. full model: AIC difference English: 52; $\chi^2(2) = 56.5$; $p < .001$; French: 6; $\chi^2(2) = 10.5$; $p = .005$; Npixel only vs. full model: AIC difference English: 24; $\chi^2(2) = 28.6$; $p < .001$; French: 3; $\chi^2(2) = 7.8$; $p = .02$). To summarize, for English and French, model comparisons showed that in addition to the prediction error, the parameter reflecting more directly the pure bottom-up processing of the physical stimulus input explained a substantial amount of variance. Nevertheless, we found that the orthographic prediction error was relevant in accounting for word recognition behavior in German, English, and French.

Word recognition behavior under conditions of visual noise

The fact that we used large-scale open source data sets for the generalization to English and French implicated a number of sources of additional of perceptual variability, which may have led to the greater role for bottom-up input in these two datasets. For example, word and font characteristics varied, e.g., word length changes from trial to trial (English, 2-13 letters; French, 2-19 letters) or the use of proportional fonts (Times new roman in the English dataset), while we had used only five-letter words presented in a monospaced font in the German experiment and the implementation of the VOP model reported here. For a predictive system, such unpredictable perceptual variation may reduce the ability to predict visual features of upcoming stimuli. For example, presenting different word lengths in a random sequence reduces the predictability of letter positions. Similarly, using a proportional-spaced font (like Times new roman) removes the letter separation in the prediction, which in turn increases the correlation between the number of pixels and the orthographic prediction error (cp. monospace font: $r = .05$ vs. proportional font: $r = .49$, both in German; see Supplement 2). The loss of structure in the sensory input, thus may result in less precise predictions and thus in larger prediction errors, which as a consequence are more similar to the total amount of bottom-up information (i.e., number of pixels) of the same word. In the face of this, it is noteworthy and important that the orthographic prediction error, as proposed here, is still highly relevant in the English and French data set, where stimuli were perceptually more variable than in the German experiment. Note, however, that in natural reading, low level visual features like word length or letter position can be picked up in parafoveal vision, so that the visual system may dynamically adapt its predictions to the upcoming word (Schotter et al., 2012).

To directly test if visual word recognition relies more strongly on the bottom-up input when visual word presentation includes unpredictable perceptual variations, we realized a second lexical decision task in German including an explicit visual noise manipulation reducing the predictability of visual features in words. We used a noise manipulation, instead of e.g. a font manipulation, since noise levels can be easily manipulated and quantified (i.e. number of displaced pixels) and a direct comparison of fonts is more difficult to realize, because the contrast of proportional vs. mono-spaced font is confounded with multiple other visual differences like total stimulus width (Hautala, Hyönä, & Aro, 2011). In addition, the 0% noise words allowed us to replicate our original behavioral finding. Figure 5a shows examples of words in monospace font with applied visual noise which were used in this experiment.

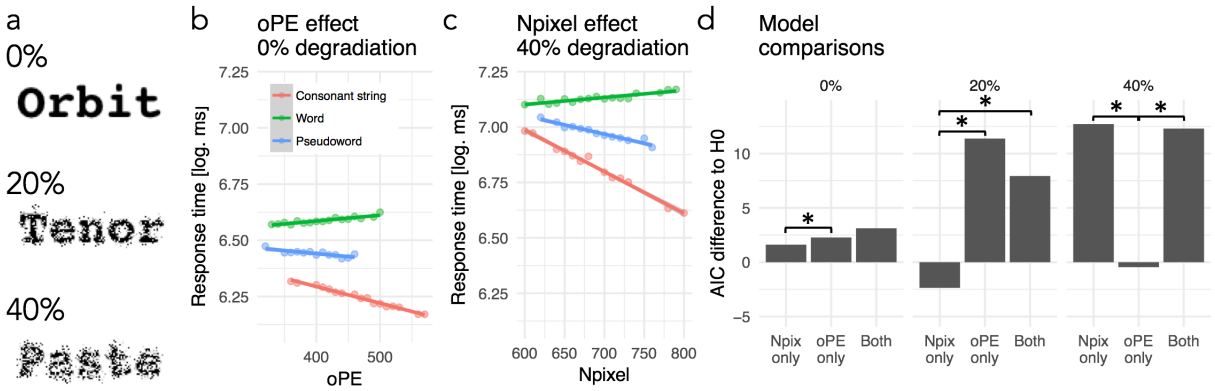


Figure 4. Stimuli and results for word/non-word decision task behavior with visual noise manipulation. (a) Example stimuli representing the three visual noise levels. (b) Orthographic prediction error effect (oPE) when no noise was applied, replicating the first study presented in Fig. 3a (error rate: 6%). (c) Number of pixels effect (Npixel) in the condition where noise was strongest (error rate: 33%). (d) Model comparisons including the full models and the models with oPE and Npixel only for each of the noise levels. Bars show the AIC difference to the null model of all nine models; asterisks mark significant differences ($p < .05$ Bonferroni corrected for multiple comparisons; 6 comparisons, three in relation to the null model and three comparing the alternative models; corrected significance threshold $p < .0083$)

We found, in general, that response times and errors increased with the amount of noise that was applied to the visual-orthographic stimuli (0%: response time/RT: 613 ms, 6% errors; 20%: RT: 739 ms, 12% errors; 40%: RT: 1,105 ms, 33% errors; compare also Fig. 4b and c). When no noise was applied we replicated our first study (cp. Fig. 4b and Fig. 3a) with a significant interaction of the orthographic prediction error and word/non-word distinction (estimate: 0.05; SE = 0.02; $t = 2.3$; see Supplemental table 1 for detailed results). No effect or

interaction was found for the number of pixels parameter. With 20% noise, we still could identify a fixed negative effect of the orthographic prediction error (estimate: -0.06; SE = 0.02; $t = 3.3$) however without a significant interaction pattern or number of pixels effect. With 40% noise, however, no significant effect of the orthographic prediction error could be found but we observed a significant fixed effect and interaction of the number of pixels parameter (Fig. 4c; estimate: 0.08; SE = 0.03; $t = 2.9$). A similar impression can be obtained from the model fit results showing that including the orthographic prediction error resulted in significantly higher model fits for 0% and 20% noise conditions compared to model were only the number of pixels predictor was included (see Fig. 4d; 0% AIC difference: 1; $\chi^2(0) = 1$; $p < .001$; 20% AIC difference: 13; $\chi^2(0) = 13.7$; $p < .001$). With 40% noise, inclusion of the number of pixels parameter resulted in a higher model fit (AIC difference: 13; $\chi^2(0) = 13.2$; $p < .001$). Surprisingly, we found an interaction between word/non-words and the number of pixels, with a positive effect for words and strong negative effects for non-words. This pattern closely matches that observed for the orthographic prediction error in stimuli without noise (cp. Fig. 4b and c).

We interpret this pattern of effects as consistent with the assumptions of predictive coding, i.e., that better predictability of the expected input (here resulting from lower visual-perceptual variability) results in more precise and stronger predictions and, therefore, greater reliance on the orthographic prediction error then on the pure bottom-up sensory input. As already discussed above, this may provide a possible explanation for the relatively increased importance of the bottom-up stimulus input in the English and French as compared to the German word recognition datasets, as in the two former the perceptual variability was higher than in the latter (due to the inclusion of stimuli of variable lengths and, in the case of the English study, proportional font). Most generally, the behavioral experiments reported in this section demonstrate that the orthographic prediction error contributes substantially to visual word recognition.

Cortical Representation of the Orthographic Prediction Error

The VOP model assumes that the orthographic prediction error is estimated at an early stage of the word recognition processes, i.e., in the visual-perceptual system and before word meaning is accessed and higher-level linguistic representations of the word can be activated.

Involved brain systems should accordingly be driven by the orthographic prediction error independent of the item's word/non-word status (i.e., for words and non-words alike). Localizing the neural signature of the orthographic prediction error in the brain during word recognition, thus, is a further critical test of the VOP model. Of note, a strict bottom-up model of word recognition (and perception in general) would make a different prediction, i.e., that activation in early visual-sensory brain regions should be driven by the full amount of physical information in the percept (Goodyear & Menon, 1998; Henrie & Shapley, 2005). Processes that take place after word identification, i.e., that involve higher levels of linguistic elaboration, can only operate on words, so that brain regions involved in these later stages of word processing should distinguish between words and non-words.

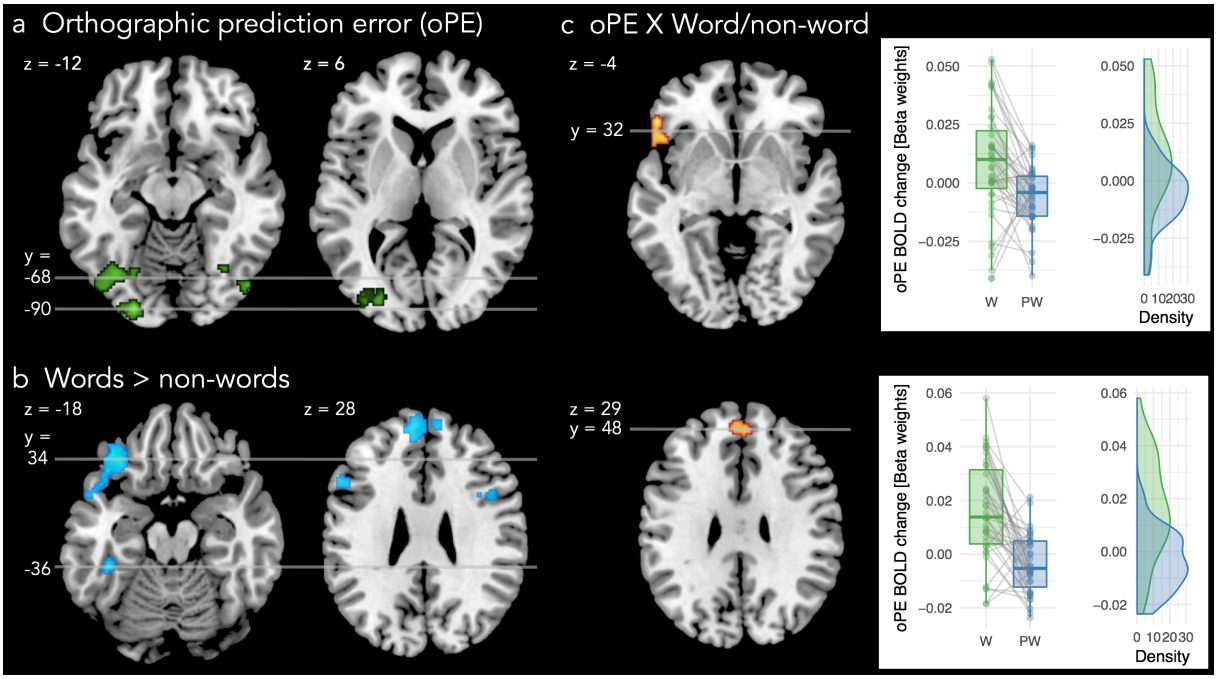


Figure 5. fMRI results demonstrating the neuroanatomical localization of orthographic prediction error effects. BOLD activation during silent reading (see Methods for further details, and Table 1 for exact locations of activation effects): (a) Analysis demonstrating a positive orthographic prediction error (oPE) effect in bilateral occipital activation-clusters. This regression analysis used item-specific oPE values as covariate, independent of stimulus condition, and shows brain regions with greater activity for letter strings characterized by a higher oPE, independent of stimulus type. (b) Clusters of higher BOLD activation for words than for non-words. (c) Two frontal activation clusters showing a oPE by word/non-word interaction, i.e. positive and negative oPE effects for words and non-words, respectively. Boxplots show individual beta weights; lines connect word and non-word betas from each individual. No effects of the number of pixels per word were found. Threshold voxel

level: p<.001 uncorrected; cluster level: p<.05 family-wise error corrected. Boxplots represent the median (line), the data from the first to the third quartile (box) and ±1.5 times the interquartile range (i.e. quartile 3 minus quartile 1; whiskers).

Table 1. Reliable activation clusters from the fMRI evaluation with respective anatomical labels (most likely regions from the Harvard-Oxford atlas; order of brain regions is relative to the order of peak components), cluster size (in voxels of size 2x2x2), and peak voxel coordinates (MNI space).

	Hemisphere	Cluster extent [N voxels]	T	x	y	z
Orthographic prediction error based analysis (positive relationship)						
Occipital fusiform gyrus / Lateral occipital gyrus	L	95	6.6*	-24	-90	-12
			4.3	-34	-88	-10
Lateral occipital gyrus	L	81	4.8	-28	-84	6
			4.3	-34	-76	6
			3.8	-38	-86	4
Lateral occipital gyrus / Occipital fusiform gyrus	R	104	5.1	48	-76	-12
			4.1	44	-64	-18
			4.0	34	-64	-14
Occipital fusiform gyrus / Lateral occipital gyrus	L	170	4.9	-36	-68	-12
			4.2	-48	-76	-10
			4.0	-24	-68	-12
Words > Pseudowords						
Frontal orbital cortex / Inferior frontal gyrus, pars triangularis	L	1347	6.6	-36	34	-18
			6.3	-40	28	-8
			6.1	-54	26	-4
Superior frontal gyrus / Frontal pole	L/R	427	5.5	-6	52	28
			3.9	-10	62	22

			3.7	10	56	26
Temporal Fusiform Cortex, posterior division	L	120	5.2	-40	-36	-18
			5.2	-34	-42	-24
Middle Temporal Gyrus, posterior division / Superior Temporal Gyrus, posterior division / Middle Temporal Gyrus, temporooccipital part	R	113	4.5	60	-34	-2
			4.0	50	-26	-2
			3.8	52	-38	0
Inferior Frontal Gyrus, pars triangularis / Frontal Pole	R	164	4.3	56	32	10
			3.9	48	34	-12
			3.4	50	34	-4
Precentral Gyrus / Inferior Frontal Gyrus, pars opercularis	R	98	3.9	44	10	28
			3.9	38	4	32
			3.7	42	16	22
Orthographic prediction error by word/non-word interaction (positive relationship for words and negative for non-words)						
Inferior frontal gyrus, pars triangularis / Frontal operculum cortex	L	125	5.5	-52	32	-4
			4.6	-48	20	-4
Paracingulate gyrus / Superior frontal gyrus	L/R	90	4.4	-4	48	28
			4.2	4	48	30

Note. Cluster-level FWE-corrected at $p < .05$, peak-level uncorrected at $p < .001$; * Significant after FWE-correction on the voxel level. Order of regions presented per cluster corresponds to the order retrieved from the probabilistic Harvard-Oxford atlas.

We tested these hypotheses about the localization of the orthographic prediction error by measuring BOLD activation changes using functional MRI while 39 participants silently read words (German nouns) and pronounceable non-words (i.e., pseudowords), in randomized order (see Methods for details). Consistent with our prediction, we identified three left- and

one right-hemispheric brain regions located in occipital cortex, that showed greater levels of activation when reading items with higher orthographic prediction error – both for words and non-words (Fig. 5a and Table 1). Importantly, no brain areas showed activity dependent on the pure amount of bottom-up information in the percept (pixel count). Prior research (Dehaene & Cohen, 2011; Dehaene et al., 2005) has identified a region in the mid-portion of the left occipito-temporal cortex as critical for reading. All four activation clusters representing the orthographic prediction error are located posterior to this so-called visual word form area (Dehaene & Cohen, 2011), which supports our claim of an early role for the orthographic prediction error signal prior to word identification.

Only brain regions involved in the activation of word meaning and subsequent processes should differentiate between words and non-words. We observed greater activity for words than non-words, independent of the orthographic prediction error, more anteriorly in left temporal and prefrontal cortex (Fig. 5b and Table 1). Third, the left inferior frontal gyrus (pars triangularis) and the medial superior frontal gyrus (mSFG) mirrored the word/non-word decision behavior reported above, in that higher prediction errors lead to increasing activation for words but decreasing activation for non-words (Fig. 5c and Table 1). The fMRI experiment, thus, supports our hypothesis that during the earliest stages of visual processing, i.e., presumably prior to accessing word meaning, an optimized perceptual signal, the orthographic prediction error, is generated and used as a basis for efficient visual-orthographic processing of written language. Only at later processing stages (in more anterior temporal and prefrontal cortices), the brain differentiates between words and non-words.

Timing of the Orthographic Prediction Error

While a representation of the orthographic prediction error could be localized in presumably 'early' visual brain regions, the temporal resolution of fMRI on the order of several seconds precludes inferences concerning the temporal sequence of cognitive processes during word recognition. The millisecond time resolution of EEG has helped to attribute the extraction of meaning-from perceived words to a time window of around 300 to 600 ms (N400 component of the event-related brain potential/ERP; Kutas & Federmeier, 2011). Visual-orthographic processes associated with the orthographic prediction error should thus temporally precede this time window, most likely to occur during the N170 component of the ERP (Barber & Kutas, 2007; Carreiras, Armstrong, Perea, & Frost, 2014). To test this hypothesis, we

measured EEG while 31 participants silently read words and non-words. A multiple regression model (analogous to the model used for the analysis of behavioral data) was fitted to the EEG data (Linzen & Engemann, 2017) with orthographic prediction error, number of pixels, word/non-word-status, and their interactions as parameters (see Methods for details).

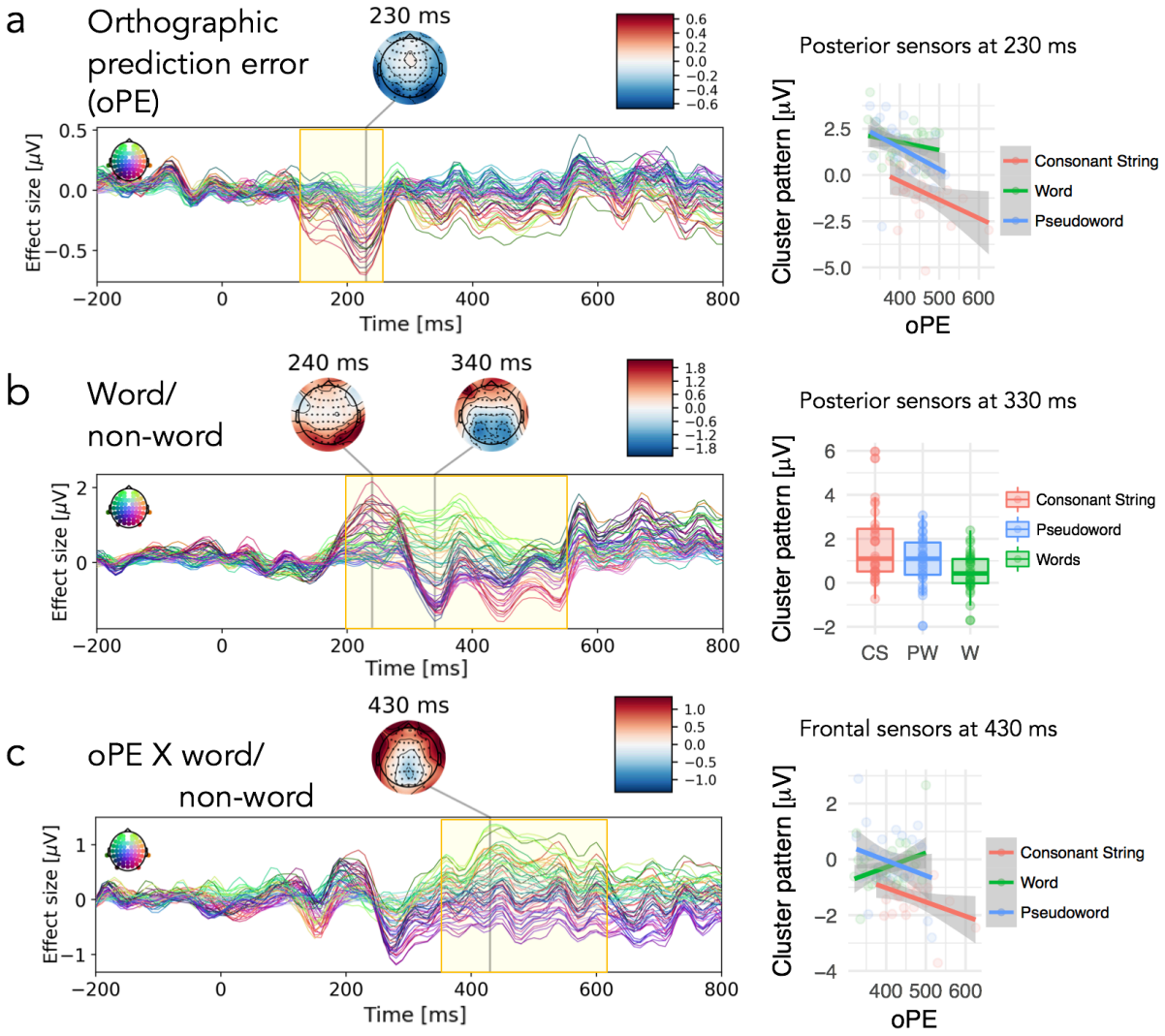


Figure 6. EEG results: Timing of orthographic prediction error effects. Effect sizes from regression ERPs are presented as time courses for each sensor and time-point (left column; color coding reflects scalp position) with yellow areas marking time windows with significant activation clusters for silent reading of 200 words and 200 non-words (100 pronouncable pseudowords, 100 consonant strings; see Supplemental figure 3 for a more detailed visualization of the significance of spatio-temporal activation clusters). ERP results are shown for (a) the orthographic prediction error (oPE) main effect, (b) the word/non-word effect, and (c) the oPE by word/non-word interaction. Results indicate significant oPE, word/non-word, and

oPE by word/non-word effects starting around, 150, 200, and 360 ms, respectively. The right panel shows the activation patterns related to the significant activation clusters (cf. Supplemental figure 3) in more detail. Dots represent mean predicted μV across (a,c) all participants and items separated by oPE and stimulus category, and (b) all items separated by stimulus category, excluding confounding effects (see Methods). No significant activation clusters were found for the parameter representing the number of pixels. Boxplots represent the median (line), the data from the first to the third quartile (box) and ±1.5 times the interquartile range (i.e. quartile 3 minus quartile 1; whiskers). The frontal cluster includes the following sensors: AF3, AF4, AF7, AF8, F1, F2, F3, F4, F5, F6, F7, F8, SO1, SO2, FP1, FP2, Fz. The posterior cluster includes the following sensors: O2, O1, Oz, PO10, PO3, PO4, PO7, PO8, PO9, POz.

Regression-estimated ERPs show a significant effect of the orthographic prediction error on electrical brain activity between 150 and 250 ms after stimulus onset (Fig. 6a). In this early time window, letter-strings characterized by higher prediction errors elicited significantly more negative-going ERPs over posterior-occipital sensors, for both words and non-words. In line with the temporal sequence of processes inferred from their neuroanatomical localizations (i.e., fMRI results), a significant word/non-word effect then emerged between 200-570 ms (Fig. 6b), followed by an interaction between word/non-word-status and orthographic prediction error at 360-620 ms (Fig. 6c). In this interaction cluster, greater prediction errors led to more negative-going ERPs for non-words, as observed for all stimuli in the earlier time window, but showed a reverse effect for words, i.e., more positive-going ERPs for words with higher prediction errors (Fig. 6c). This pattern of opposite prediction error effects for words vs. non-words is analogous to the effects seen in word/non-word decision behavior and in the frontal brain activation results.

As in the fMRI study, we found no effect of the bottom-up input as such (pixel count), even though it is well-established that manipulations of physical input contrast (as determined, e.g., by the strength of luminance; Johannes, Münte, Heinze, & Mangun, 1995) can increase the amplitude of early ERP components starting at around 100 ms. We performed an explicit model comparison between statistical models including the orthographic prediction error compared to a model including the number of pixels parameter (analogous to the analysis of behavioral data), for both time windows in which the orthographic prediction error was relevant (early fixed effect and later interaction). In both time windows the model including the orthographic prediction error resulted in better fit (AIC

difference: 16 at 230 ms at posterior sensors: $\chi^2(0) = 16.0$; $p < .001$; and 5 at 430 ms at frontal sensors: $\chi^2(0) = 5.0$; $p < .001$).

To summarize, EEG results converge with fMRI results and suggest that relatively early on in the cortical visual-perceptual processing cascade, the amount of perceptual processing devoted to the orthographic percept is smallest for letter-strings with highly expected visual features (i.e., low orthographic prediction error). 100 to 200 ms later, i.e., in a time window strongly associated with semantic processing (Kutas & Federmeier, 2011), the prediction error effect was selectively reversed for words, and thus started to differentiate between the two stimulus categories. This mirrors behavioral results and activation patterns in anterior temporal lobe and prefrontal cortex. Combined, these results support the VOP model's proposal that orthographic representations are optimized early during visual word recognition, and that the resulting orthographic prediction error is the basis for subsequent stages of word recognition.

Applying the VOP model to handwritten script.

The electronic fonts used for all above-reported experiments introduce a highly regular structure that favors some of the VOP model's core processes, like the calculation of the orthographic prediction error. We showed above that reducing the high regularity of computerized script by visual noisy, reading performance decreases and the orthographic prediction error becomes less relevant for describing reading behavior. To demonstrate the 'real world' validity of the VOP model with even less regular scripts, we applied a variant of this model to naturalistic reading of handwriting. The extreme variability of different handwritings strongly influences their readability. Visual-orthographic predictions, here implemented on the basis of single letters and separately for each handwriting, accordingly vary substantially in the strength and precision between individual handwritings (cp. prediction of Fig. 7a,b). Prediction strength is represented in terms of darkness of the gray values of the prediction image, i.e., the mean gray value across pixels. The precision of the prediction is represented by the inverse of the number of gray pixels included the prediction image; more precise predictions are more focused and less distributed. High strength and high precision in the script-specific prediction result in lower orthographic prediction errors (Fig. 7c,d; linear mixed model statistics: Estimate: -0.05; SE = 0.01; $t = 7.4$ and estimate: 0.02; SE

= 0.01; t = 2.1, respectively; see Supplemental table 1 for full results). Finally, we obtained the rated readability of the handwriting on the basis of 10 handwritten words and observed that the readability is higher for handwritings that produce lower prediction errors (Fig. 7e; 38 raters; Estimate: -5.9; SE = 1.0; t = 6.2). This variant of the VOP model, shows that we can account for reading processes not only in highly formalized stimuli, but also when applied to handwritings.

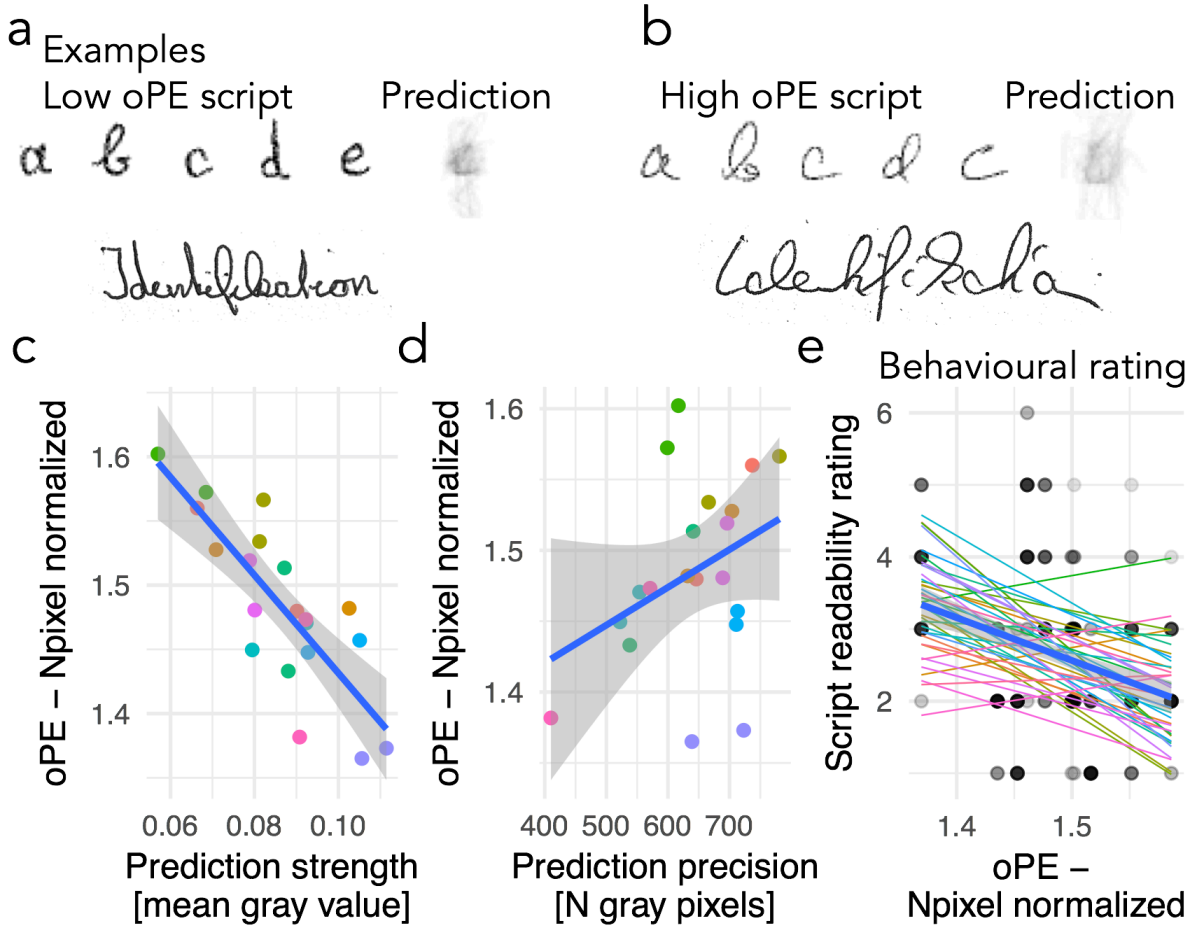


Figure 7. Applying the visual-orthographic prediction model of reading to the perception of handwritings. Examples for two (out of 10 empirically obtained) different handwritings. (a) A handwriting including single letters and the respective (letter-level) orthographic prediction estimated based on all 26 lower case letters (written in isolation). In addition, the word Identifikation (identification) is presented for both handwritings, as an example (out of 10). These words were used to acquire the readability rating. (c) Relationship between prediction strength and the mean orthographic prediction error across all letters for each script. Note that the oPE estimate for handwritings was normalized (i.e. divided) by the number of pixels since the number of pixels differed drastically between scripts (e.g. compare Examples in a and b). (d) Relationship between the precision of the prediction and the orthographic prediction error. Point color reflects each of 10 individual scripts,

separately for upper- and lower-case letters. (e) Script readability ratings in relation to the orthographic prediction error (lower- and upper-case prediction error combined). Blue line reflects the overall relationship and thin lines represent each rater.

Discussion

We have demonstrated that predictive coding is a plausible model for visual-orthographic processing during reading. This conclusion is empirically supported by our observation that the orthographic prediction error, i.e., the non-redundant and thus informative part of the visual-orthographic percept, (i) accounts for word identification behavior, (ii) explains brain activation in low-level visual-perceptual systems of the occipital cortex, and (iii) explains brain activation as early as 150 ms after the onset of the letter-string. In addition, our visual orthographic prediction (VOP) model provides a quantitative estimate of the amount of information reduction achieved by this mechanism (i.e., in our data between 29 and 37% on average depending on language, with an upper limit of 51% at the level of the individual word). The pattern of differential correlations with various lexicon-based descriptors of words supports our proposal of an association of this prediction error representation with orthographic stages of visual word recognition. Finally, we have provided first evidence that the basic ideas of the VOP model may also be applicable to more naturalistic reading situations, for example to account for individual differences in the readability of handwritings. Combined, these findings are evidence against the common (explicit or implicit) assumption that the pure bottom-up visual information (i.e., in terms of the total number of pixels included in a word) is the basis for the transformation of visual input into abstract representations of letters and words that is at the core of the reading process. Rather, our results suggest that top-down guided predictive processing is used to optimize the visual input and improve the efficiency of word recognition, even when reading isolated words.

We also found evidence for a greater reliance on bottom-up visual information when higher levels of variability are present in the visual stimulus – i.e., when visual occurrence of the stimulus is less predictable, for example due to visual noise or a larger range of different word lengths. Already some of the earliest models of predictive coding (Rao & Ballard, 1999) showed that noise in the percept reduces the amount of predictable information. A shift to bottom-up processing, thus, may represent an important fallback strategy of predictive systems. In naturalistic sentence or text reading, the perceptual variability introduced by

factors such as variable word lengths are accounted for by the fact that low-level visual properties (e.g., letter length) are available prior to fixation through para-foveal vision (Schotter et al., 2012). The visual system, accordingly, in principle has the capability of dynamically implementing best-fitting visual-orthographic predictions, thereby allowing optimized sensory processing as described by the VOP model in most natural reading situations. This hypothesis must be tested in future studies but fits with proposals which have acknowledged the integration of top-down predictions from multiple linguistic domains (for example at the phonological, semantic, or syntactic level; DeLong et al., 2005; Nieuwland et al., 2018; Price & Devlin, 2011). Critically, our results go clearly beyond this by demonstrating that top-down guided expectations are implemented already onto early visual-perceptual processing stages.

The so-far dominant model of visual word recognition in the brain (Dehaene & Cohen, 2011; Dehaene et al., 2005) postulates that words are 'assembled' in a bottom-up fashion from line- and orientation-sensitive receptive fields of simple primary visual cortex neurons into successively more complex higher-order representations along the visual pathway. The high correlation between the representational spaces of our original word stimulus images and their derived orthographic prediction errors (Fig. 2d), however, indicates that sufficient information is retained in the optimized input representation so that words can be discriminated with high precision. As a consequence, extraction of feature combinations (like letters or bigrams; cf. Dehaene et al., 2005) should in principle be possible, however according to our account in a more efficient way from the orthographic prediction error rather than from a full representation of the bottom-up input. Prediction-based top-down optimization of the visual-orthographic input, as proposed here, is thus not necessarily incompatible with the currently prevalent bottom-up models of word recognition. Nevertheless, the two theoretical accounts are fundamentally different, for example with respect to the role of top-down processing. Future research will have to show which theory can better account for neuronal processing at various stages of the hierarchical word processing system.

In sum, we demonstrate that during reading, visual-orthographic information processing is optimized by explaining away redundant visual information. This provides strong evidence that reading follows domain-general mechanisms of predictive coding during perception (Clark, 2013) and is also consistent with the influential hypothesis of a Bayesian brain, which during perception continuously combines prior knowledge and new sensory evidence (K. Friston, 2005; Knill & Pouget, 2004). We propose that the result of this

optimization step, i.e., an orthographic prediction error signal, is the neurophysiologically efficient access code to subsequent higher levels of word processing, including the activation of word meaning. These data provide the basis for a new understanding of visual word recognition rooted in domain-general neurophysiological mechanisms of prediction-based perceptual processes (K. Friston, 2005; Rao & Ballard, 1999). At the same time, our results provide important converging evidence in support of predictive coding theory.

Methods

Implementation of the VOP Model

The estimation of the orthographic prediction error as assumed in VOP was realized by image-based computations. Using the *EBImage* package in *R* (Pau, Fuchs, Sklyar, Boutros, & Huber, 2010), letter-strings were transformed into gray scale images (size for, e.g., 5-letter words: 140x40 pixels) that can be represented by a 2-dimensional matrix in which white is represented as 1, black as 0, and gray as intermediate values. This matrix representation allows an easy implementation of the subtraction computation presented in Fig. 1a, i.e.,

$$(1)\begin{bmatrix} SI_{1,1} & \dots & SI_{140,1} \\ \vdots & \ddots & \vdots \\ SI_{1,40} & \dots & SI_{140,40} \end{bmatrix} - \begin{bmatrix} P_{1,1} & \dots & P_{140,1} \\ \vdots & \ddots & \vdots \\ P_{1,40} & \dots & P_{140,40} \end{bmatrix} = \begin{bmatrix} oPE_{1,1} & \dots & oPE_{140,1} \\ \vdots & \ddots & \vdots \\ oPE_{1,40} & \dots & oPE_{140,40} \end{bmatrix}$$

where $SI_{x,y}$ indicates the sensory input at each pixel. $P_{x,y}$ reflects the prediction matrix which is in the present study calculated as an average across all words (or a subset thereof) in a lexical database e.g., the example shown in Fig. 1b is based on 5,896 nouns of five letters length from the English SUBTLEX database (Heuven et al., 2014). This orthographic prediction was estimated by transforming each of n words into a matrix as described above and then averaging the values included in these matrices:

$$(2)\frac{\sum_1^n \begin{bmatrix} SI_{1,1} & \dots & SI_{140,1} \\ \vdots & \ddots & \vdots \\ SI_{1,40} & \dots & SI_{140,40} \end{bmatrix}}{n} = \begin{bmatrix} P_{1,1} & \dots & P_{140,1} \\ \vdots & \ddots & \vdots \\ P_{1,40} & \dots & P_{140,40} \end{bmatrix}$$

The VOP model postulates that during word processing, *SI* is reduced by the prediction matrix *P*, resulting in an orthographic prediction error matrix (*oPE*) as shown above in formula (1). The resulting orthographic predication error is therefore black (i.e. value 0) at pixels were the

prediction was perfect and gray to white (i.e. value > 0) where the visual information was not predicted perfectly. As a last step, a numeric value for the orthographic prediction error of each stimulus was determined by summing all values of its prediction error matrix. This numeric representation of the prediction error is used as parameter for all empirical evaluations.

$$(3)\sum\begin{bmatrix} oPE_{1,1} & \dots & oPE_{140,1} \\ \vdots & \ddots & \vdots \\ oPE_{1,40} & \dots & oPE_{140,40} \end{bmatrix} = oPE_{sum}$$

The amount of information reduction ($I_{reduced}$) achieved by this predictive computation can then be calculated by relating the numeric representation of the prediction error to an analogous numeric representation of the respective word SIsum:

$$(4)1 - \frac{oPE_{sum}}{SI_{sum}} * 100 = I_{reduced}$$

Participants

35, 54, 39, 31, and 38 healthy volunteers (age from 18 to 39) participated in the two German lexical decision studies, the fMRI, the EEG, and the handwriting experiments, respectively. All had normal reading speed (reading scores above 20th percentile estimated by a standardized screening; unpublished adult version of Auer, Guber, Wimmer, & Mayringer, 2005), reported absence of speech difficulties, had no history of neurological diseases, and normal or corrected-to-normal vision. Participants gave written informed consent and received student credit or financial compensation (10€/h) as incentive for participating. The research was approved by the ethics board of the University of Salzburg (EK-GZ: 20/2014; fMRI study) and Goethe University Frankfurt (#2015-229; EEG study, lexical decision studies). Behavioral results for English, and French were obtained from publicly available data sets, whose samples are described elsewhere (Ferrand et al., 2010; Keuleers et al., 2012).

Materials, experimental procedures, and analyses.

Lexicon-based Characterization of the Orthographic Prediction Error. We calculated the number of pixels per word, the orthographic prediction error, and established word characteristics (Orthographic Levenshtein distance; Yarkoni et al., 2008, word frequency) for 3,110 German (Brysbaert et al., 2011) nouns (i.e., the subset used for the empirical evaluations later on; with uppercase first letters), for 5,896 English (Heuven et al., 2014)

words, 5,638 French (New, Pallier, Brysbaert, & Ferrand, 2004) words, and 4,418 Dutch (Keuleers, Brysbaert, et al., 2010) words. All items had a length of five letters. For the German nouns, we additionally estimated a more comprehensive set of orthographic word characteristics, including bi-, tri-, quadirgram-frequencies (i.e., occurrences of 2, 3, 4 letter combinations), and Coltheart's N (Coltheart et al., 1977); see Fig. 2b). Orthographic Levenshtein distance and Coltheart's N were estimated with the *vwr* Package in R (Keuleers, 2013).

Accounting for Word Recognition Behavior. German lexical decision task 1: 800 five-letter nouns and 800 five-letter nonwords (400 pronouncable pseudowords, 400 unpronouncable non-words/consonant clusters) were presented in pseudorandomized order (Experiment Builder software, SR-Research, Ontario, Canada; black on white background; Courier-New font; .3° of visual angle per letter; 21″ LCD monitor with 1,024 × 768 resolution and 60Hz refresh rate), preceded by 10 practice trials. Participants judged for each letter string whether it was a word or not using a regular PC keyboard, with left and right arrow keys for words and non-words, respectively. Before stimulus presentation, two black vertical bars (one above and one below the vertical position of the letter string) were presented for 500 ms, and letter strings were displayed until a button was pressed. Response times were measured in relation to the stimulus onset. German lexical decision task 2 including noisy stimuli reports a replication in German with 70 five-letter words and 70 nonwords (36 pseudowords, 34 consonant clusters) with no noise with identical procedures except that data were acquired in small groups of up to 8 participants. In addition, words with 20% or 40% noise added (i.e. 20% or 40% of pixels were displaced; for details see Gagl et al., 2014) were presented in blocks of 140 (70 five-letter words and 70 nonwords).

Linear mixed model (LMM) analysis implemented in the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) of the R statistics software were used for analyzing lexical decision data as LMMs are optimized for estimating statistical models with crossed random effects for items. These analyses result in effect size estimates with confidence intervals (SE) and a *t*-value. Following standard procedures, *t*-values larger than 2 are considered significant since this indicates that the effect size ±2 SE does not include zero (Kliegl, Wei, Dambacher, Yan, & Zhou, 2011). For the presentation in Fig. 3a,b,g,h,j,g and 4b,c co-varying effects were removed by the *keepef* function of the *remef* package (Hohenstein & Kliegl, 2014/2017). All response times were log-transformed, which accounts for the ex-Gaussian distribution of response times. In addition, orthographic prediction error, and number of pixels were centered and normalized by R's *scale()* function in order to optimize LMM analysis.

Cortical Representation of the Orthographic Prediction Error. 60 five-letter words and 180 pseudowords were presented in pseudorandom order (yellow Courier New font on gray background; 800 ms per stimulus; ISI 2,150 ms) as well as 30 catch trials consisting of the German word *Taste* (button), indicating participants to press the response button. Catch trials were excluded from the analyses. All items consisted of two syllables and were matched on OLD20 (Yarkoni et al., 2008) and mean bigram frequency between conditions. To facilitate estimation of the hemodynamic response, an asynchrony between the TR (2,250 ms) and stimulus presentation (onset asynchrony: 2,150 + 800 ms) was established and 60 null events were interspersed among trials; a fixation cross was shown during inter-stimulus intervals and null events. The sequence of presentation was determined by a genetic algorithm (Wager & Nichols, 2003), which optimized for maximal statistical power and psychological validity. The fMRI session was divided into 2 runs with a duration of approximately 8 min each.

A Siemens Magnetom TRIO 3-Tesla scanner (Siemens AG, Erlangen, Germany) equipped with a 32-channel head-coil was used for functional and anatomical image acquisition. The BOLD signal was acquired with a T_2*-weighted gradient echo echo-planar imaging sequence (TR = 2,250 ms; TE = 30 ms; Flip angle = 70°; 86 x 86 matrix; FoV = 192 mm). Thirty-six descending axial slices with a slice thickness of 3 mm and a slice gap of 0.3 mm were acquired within each TR. In addition, for each participant a gradient echo field map (TR = 488 ms; TE 1 = 4.49 ms; TE 2 = 6.95 ms) and a high-resolution structural scan (T_1-weighted MPRAGE sequence; 1 x 1 x 1.2 mm) were acquired. Stimuli were presented using an MR-compatible LCD screen (NordicNeuroLab, Bergen, Norway) with a refresh rate of 60 Hz and a resolution of 1,024x768 pixels.

SPM8 software (http://www.fil.ion.ucl.ac.uk/spm), running on Matlab 7.6 (Mathworks, Inc., MA, USA), was used for preprocessing and statistical analysis. Functional images were realigned, unwarped, corrected for geometric distortions by use of the FieldMap toolbox, and slice-time corrected. The high-resolution structural image was pre-processed and normalized using the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm8). The image was segmented into gray matter, white matter, and CSF compartments, denoised, and warped into MNI space by registering it to the DARTEL template of the VBM8 toolbox using the high-dimensional DARTEL registration algorithm (Ashburner, 2007). Functional images were co-registered to the high-resolution structural image, which was normalized to the MNI T_1 template image, and resulting normalization parameters were applied to the functional data, which were then resampled to a resolution of 2×2×2 mm and smoothed with a 6 mm FWHM Gaussian kernel.

For statistical analysis, we first modeled stimulus onsets with a canonical hemodynamic response function and its temporal derivative, including movement parameters from the realignment step and catch trials as covariates of no interest, a high-pass filter with a cut off of 128 s, and an AR(1) model (K. J. Friston et al., 2002) to correct for autocorrelation. For the group level statistics, t-tests were realized with a voxel level threshold of $p < .001$ uncorrected and a cluster level correction for multiple comparisons ($p < .05$ family-wise error corrected).

Timing of the Orthographic Prediction Error. 200 five-letter words, 100 pseudowords, and 100 consonant strings (nonwords) were presented for 800 ms (black on white background; Courier-New font, .3° of visual angle per letter), followed by an 800 ms blank screen and a 1,500 ms hash mark presentation, which marked an interval in which the participants were instructed to blink if necessary. In addition, 60 catch trials (procedure as described for fMRI study) were included in the experiment. Stimuli were presented on a 19″ CRT monitor (resolution 1,024 × 768 pixels, refresh rate 150Hz), and were preceded by two black vertical bars presented for 500 - 1,000 ms to reduce stimulus onset expectancies.

EEG was recorded from 64 active Ag/Ag-Cl electrodes (extended 10-20 system) using an actiCAP system (BrainProducts, Germany). FCz served as common reference and the EOG was recorded from the outer canthus of each eye as well as from below the left eye. A 64-channel Brainamp (BrainProducts, Germany) amplifier with a 0.1–1,000 Hz band pass filter sampled the amplified signal with 500Hz. Electrode impedances were kept below 5kΩ. Offline, the EEG data were re-referenced to the average of all channels. EEG data were preprocessed using MNE-Python (Gramfort et al., 2014), including high (.1 Hz) and low pass (30 Hz) filtering and removal of ocular artifacts using ICA (Delorme, Sejnowski, & Makeig, 2007). For each subject, epochs from 0.5 s before to 0.8 s after word onset were extracted and baselined by subtracting the pre-stimulus mean, after rejecting trials with extreme (>50 μV peak-to-peak variation) values. Multiple regression analysis, with the exact same parameters as for the behavioral evaluation (orthographic prediction error, number of pixels, word/non-word, and the interactions with the word/non-word distinction), was conducted and a cluster-based permutation test (Maris & Oostenveld, 2007) was used for significance testing. 1,024 label permutations were conducted to estimate the distribution of thresholded clusters of spatially and temporally (i.e., across electrodes and time) adjacent time points under the null hypothesis. All clusters with a probability of less than an assumed alpha value of 0.05 under this simulated null hypothesis were considered statistically significant. The presentation of effect patterns (line

and box-plots) in Fig. 6 co-varying effects were removed by the *keepef* function of the *remef* package (Hohenstein & Kliegl, 2014/2017).

Application to handwriting. We obtained handwriting samples (26 upper and 26 lower case letters; 10 common German compound words, 10-24 letters long) from 10 different writers (see Fig. 5a,b for examples). The single letters were scanned and centered within a 50x50 pixels image. These images were used to estimate, for each script separately, pixel-by-pixel predictions for upper and lower-case letters (see also Fig. 5a,b), analogous to the procedures described above and in Fig. 1b. Subsequently, these predictions were subtracted from each letter of the alphabet, within the respective script sample (matrix subtraction; Formula 1). In contrast to computer fonts the correlation of the orthographic prediction error and the respective item's number of pixels was high (r = .98). To compensate this, the orthographic prediction error was normalized by a division with the respective pixel count. Readability ratings (5-point Likert scale) were obtained from 38 participants (27 females; mean age 25 years) by presenting all ten versions of all ten handwritten compound words, in addition to the identical word in computerized script. Note that all stimuli including lexical characteristics and original handwritings will be available on Zenodo (after acceptance).

For the handwriting data, a LMM was realized that predicted the orthographic prediction error (Fig. 5c-d) from the following parameters: mean prediction strength (i.e., mean of the values extracted from the prediction matrix), number of all non-white pixels (both scaled), and letter case. The random effect on the intercept was estimated for each script. In addition, a second LMM was estimated for readability ratings with the orthographic prediction error as the only predictor and participants as random effect on the intercept and as random effect of the orthographic prediction error slope.

Data and Code availability

Data and analysis scripts will be made publicly available at Zenodo when published.

Acknowledgements

We thank Rebekka Tenderra, Anne Hoffmann, Jan Jürges, and Kirsten Hilger for help with EEG data acquisition. In addition, we thank Mark D'Esposito, David Poeppel, Matt Davis, and Ulrike Basten for helpful comments on a previous version of the manuscript. The

research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2013) under grant agreement n° 617891 awarded to CJF and from the European Community's Horizon 2020 Programme (2016) under grant agreement n° 707932 awarded to BG.

Author Contributions

B.G. and C.F. wrote and revised the manuscript and all of the authors edited the final drafts. B.G., J.S., and C.F. conceptualized the model. B.G. and S.H. implemented the model. B.G. realized the model simulations, conceptualized behavioral experiments, and behavioral data analysis. B.G. and K.G. realized behavioral data acquisition, and the handwriting adaptation. B.G., J.S., and S.H. designed, measured and analyzed the EEG study. B.G. and F.R. designed, measured and analyzed the fMRI study.

References

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801. https://doi.org/10.1038/nn.2810

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Auer, M., Guber, G., Wimmer, H., & Mayringer, H. (2005). *Salzburger Lese-Screening für die Klassenstufen 1-4*. Hogrefe, Verlag für Psychologie. Retrieved from https://www.testzentrale.de/shop/salzburger-lese-screening-fuer-die-klassenstufen-1-4.html

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology. General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, *53*(1), 98–123. https://doi.org/10.1016/j.brainresrev.2006.07.002

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The Word Frequency Effect. *Experimental Psychology*, *58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458. https://doi.org/10.1037/xhp0000159

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Changizi, M. A., Zhang, Q., Ye, H., & Shimojo, S. (2006). The Structures of Letters and Symbols throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes. *The American Naturalist*, *167*(5), E117–E139. https://doi.org/10.1086/502806

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In *Attention and performance VI. Proceedings of the Sixth International Symposium on Attention and Performance, Stockholm, Sweden, July 28-August 1, 1975*.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1778–1786. https://doi.org/10.1037/xlm0000013

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in*

Cognitive Sciences, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, *9*(7), 335–341. https://doi.org/10.1016/j.tics.2005.05.004

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage*, *34*(4), 1443–1449.

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, *21*(4), 449–467.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., … Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. https://doi.org/10.3758/BRM.42.2.488

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian Inference in Neuroimaging: Applications. *NeuroImage*, *16*(2), 484–512. https://doi.org/10.1006/nimg.2002.1091

Gagl, B., Golch, J., Hawelka, S., Sassenhagen, J., Poeppel, D., & Fiebach, C. J. (2018). Reading at the speed of speech: the rate of eye movements aligns with auditory language processing. *BioRxiv*, 391896. https://doi.org/10.1101/391896

Gagl, B., Hawelka, S., Richlan, F., Schuster, S., & Hutzler, F. (2014). Parafoveal preprocessing in reading revisited: Evidence from a novel preview manipulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 588–595. https://doi.org/10.1037/a0034408

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal Predictive Codes for Spoken Words in Auditory Cortex. *Current Biology*, *22*(7), 615–621. https://doi.org/10.1016/j.cub.2012.02.015

Goodyear, B. G., & Menon, R. S. (1998). Effect of Luminance Contrast on BOLD fMRI Response in Human Primary Visual Areas. *Journal of Neurophysiology*, *79*(4), 2204–2207. https://doi.org/10.1152/jn.1998.79.4.2204

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*, *103*(3), 518.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Haarmann, H. (2007, March 15). Geschichte der Schrift | Haarmann, Harald | Verlag C.H.BECK Literatur - Sachbuch - Wissenschaft. Retrieved March 27, 2017, from http://www.chbeck.de/Haarmann-Geschichte-Schrift/productview.aspx?product=20173

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research*, *51*(15), 1719–1727. https://doi.org/10.1016/j.visres.2011.05.015

Henrie, J. A., & Shapley, R. (2005). LFP Power Spectra in V1 Cortex: The Graded Effect of Stimulus Contrast. *Journal of Neurophysiology*, *94*(1), 479–490. https://doi.org/10.1152/jn.00919.2004

Heuven, W. J. B. van, Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Hohenstein, S., & Kliegl, R. (2017). *remef: Remove Partial Effects*. R. Retrieved from https://github.com/hohenstein/remef (Original work published 2014)

Johannes, S., Münte, T. F., Heinze, H. J., & Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. *Cognitive Brain Research*, *2*(3), 189–205. https://doi.org/10.1016/0926-6410(95)90008-X

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Keuleers, E. (2013). vwr: Useful functions for visual word recognition research (Version 0.3.0). Retrieved from https://cran.r-project.org/web/packages/vwr/index.html

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice Effects in Large-Scale Visual Word Recognition Studies: A Lexical Decision Study on 14,000 Dutch Mono- and Disyllabic Words and Nonwords. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*, 238. https://doi.org/10.3389/fpsyg.2010.00238

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Linzen, T., & Engemann, D. (2017). Sensor space least squares regression — MNE 0.15.dev0 documentation. Retrieved June 23, 2017, from http://martinos.org/mne/dev/auto_examples/stats/plot_sensor_regression.html

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2 : A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524. https://doi.org/10.3758/BF03195598

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018, April 3). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. https://doi.org/10.7554/eLife.33468

Niven, J. E., & Laughlin, S. B. (2008). Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology*, (211), 1792–1804. https://doi.org/10.1242/jeb.017574

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., & Huber, W. (2010). EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, *26*(7), 979–981. https://doi.org/10.1093/bioinformatics/btq046

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, *15*(6), 246–253. https://doi.org/10.1016/j.tics.2011.04.001

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*(1), 5–35.

Schwiedrzik, C. M., & Freiwald, W. A. (2017). High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. *Neuron*, *96*(1), 89-97.e4. https://doi.org/10.1016/j.neuron.2017.09.007

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive Coding: A Fresh View of Inhibition in the Retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205), 427–459.

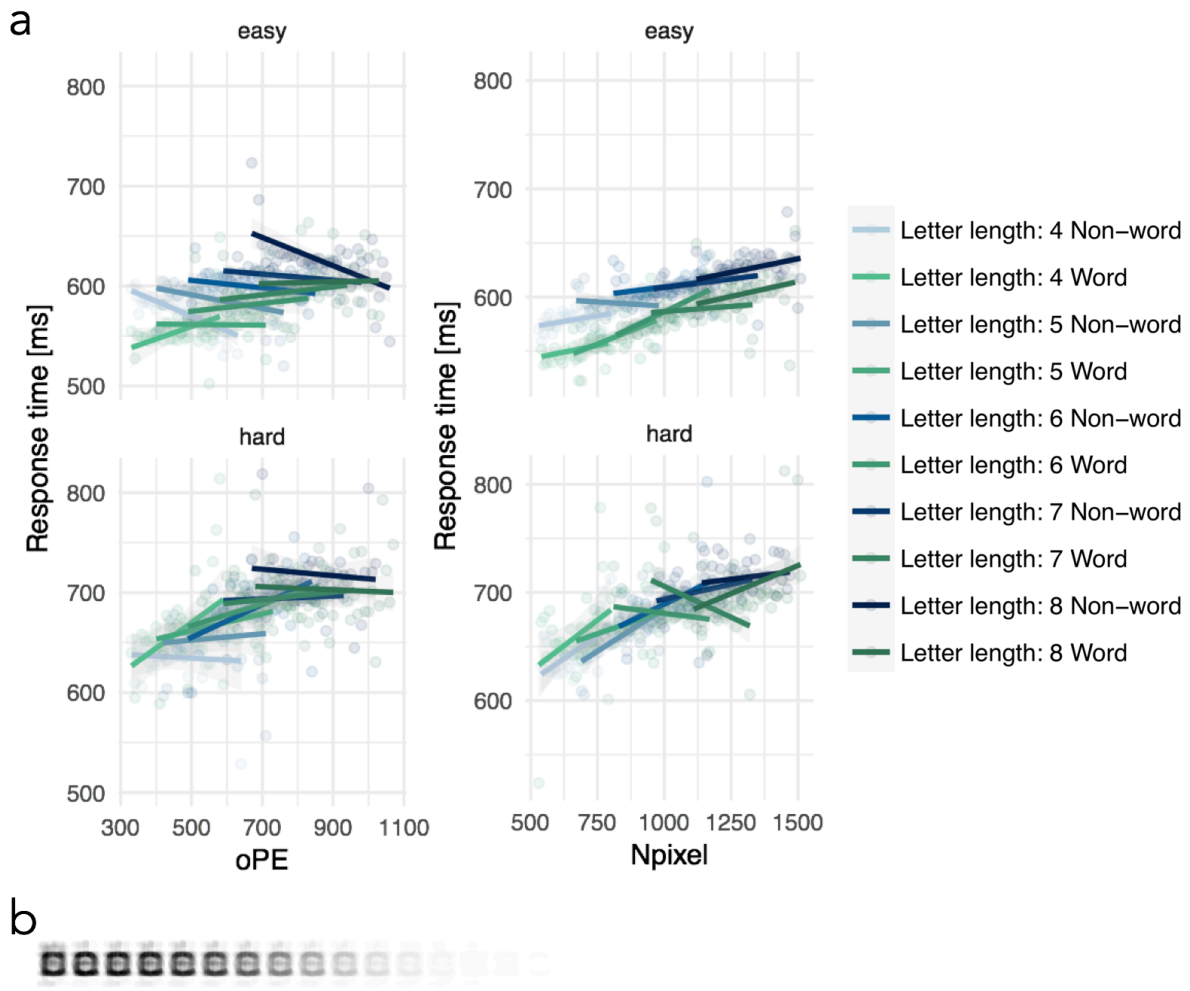
Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. *The Journal of Neuroscience*, *31*(25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage*, *18*(2), 293–309. https://doi.org/10.1016/S1053-8119(02)00046-0

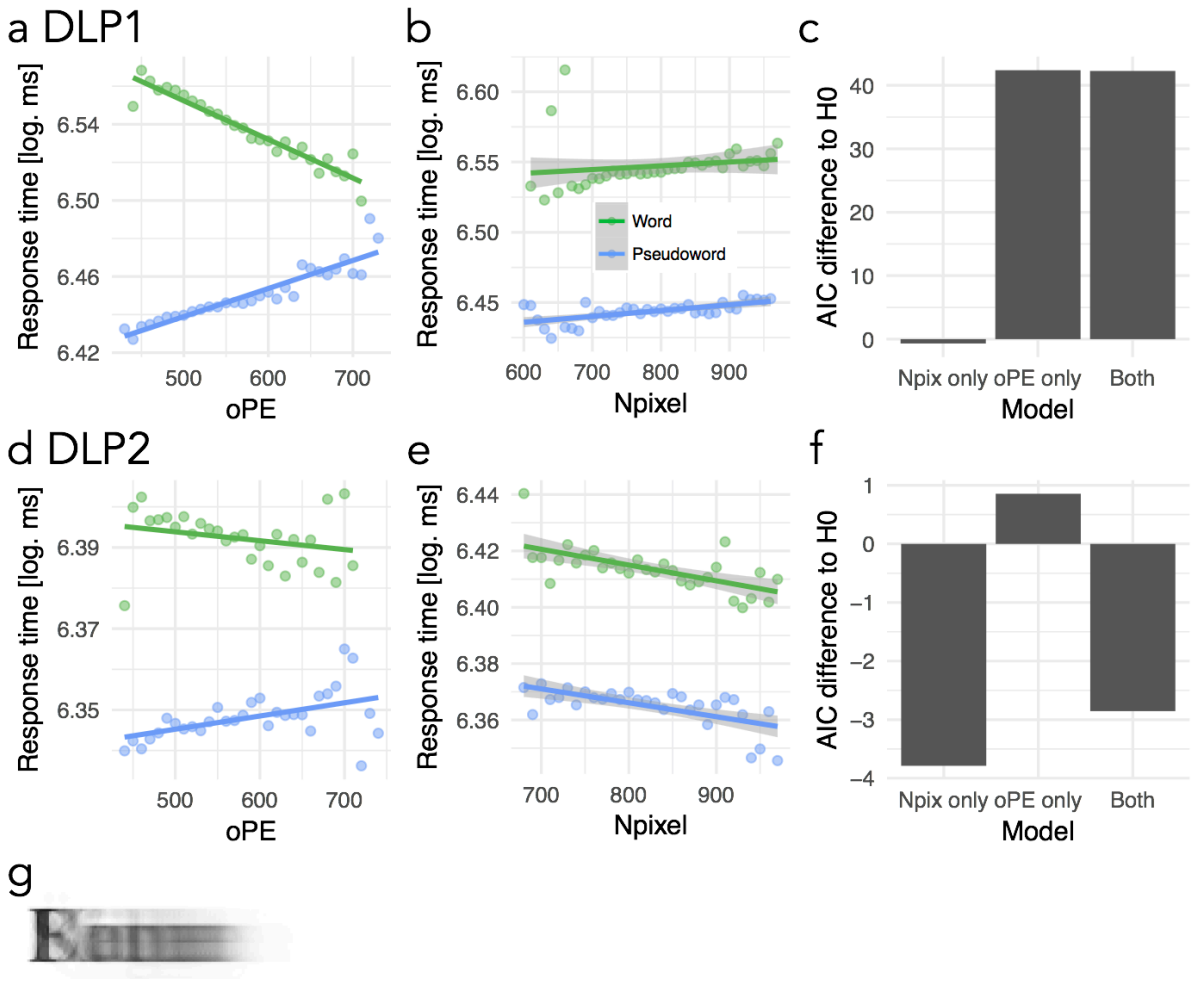
Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Supplements



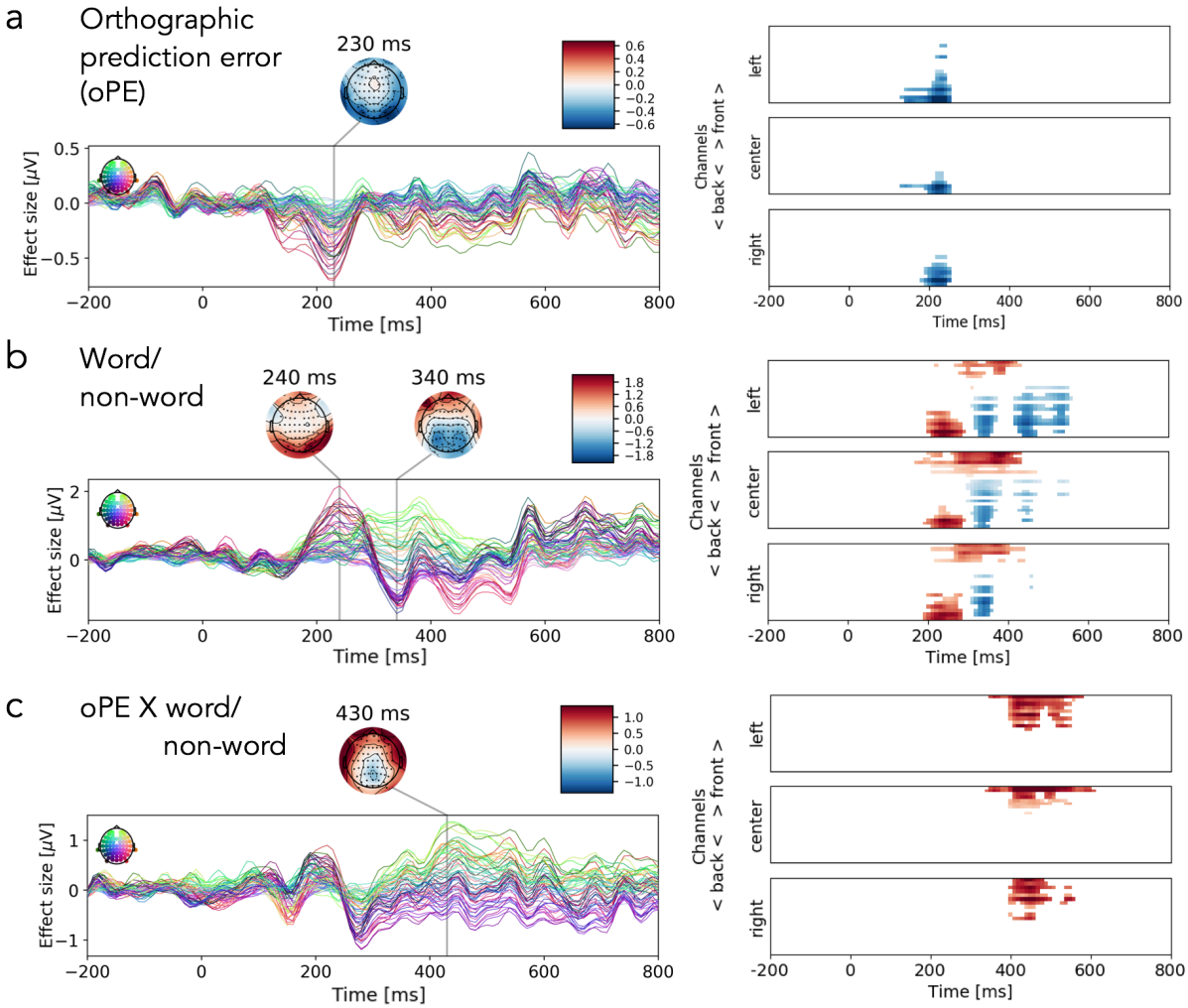
Supplemental figure 1. (a) Response times aggregated across participants from the British lexicon (BLP) project (Keuleers et al., 2012) for the word lengths 4-8. The left panel shows the word/non-word by orthographic prediction error (oPE) interaction and the right panel shows the word/non-word by number of pixels (Npixel) interaction for each word length separately. In addition, the upper panel shows letter strings that are correctly categorized in nearly all cases (accuracy > .95) and the lower panel shows the response times to the items, which were less accurately processed (i.e., accuracy < .95). We realized this median split in order to result in a subset of the BLP (i.e., the easy words) which are roughly comparable to the previous experiments (e.g. see Fig. 2d), as the BLP study includes a large number of very rare words (median log. word frequency per million is .3). Analogous to Supplemental figure 1, bluish colors represent non-words (N) and greenish colors represent words (W), while the hue of the colors reflects word length (i.e., bright to dark reflects short to long letter strings). For both effects, we first estimated linear regression models with either the oPE or the Npixel effect and allowing interactions with word/non-word status, word length, and accuracy. Note that the oPE in this first analysis was based on length-specific predictions (i.e., for the estimation of the oPE

of four-letter words, all four-letter words of the lexicon were included in the prediction). For the oPE model, a significant four-way interaction was found (estimate = -1.078e-04; SE = 4.199e-05; t = -2.567). Separating hard vs. easy words allowed us to disentangle the four-way interaction: In easy words/non-words, we found a consistent (i.e., across length levels) oPE by word/non-word interaction (estimate = 1.530e-04; SE = 4.047e-05; t = 3.780) in the same direction as previously shown (positive effect for words and a negative effect for non-words). For hard words/non-words, we found that the oPE by word/non-word interaction was inconsistent across letter length levels, which was indicated by a significant oPE and letter length interaction (estimate = -3.530e-05; SE = 8.092e-06; t = -4.363). In addition, for the hard words both the oPE by word/non-word interaction (estimate = -1.685e-04; SE = 6.905e-05; t = -2.440) and the main effect of oPE were reversed (estimate = 2.828e-04; SE = 5.802e-05; t = 4.874 compare to estimate = -1.000e-04; SE = 2.440e-05; t = -4.101, for easy words). For the Npixel model, no four-way interaction and no Npixel interaction or main effect were found. In sum, in this analysis we showed that the oPE by word/non-word interaction shown previously for word lengths of five letters (see main text) is consistent for easy-to-process English items with word lengths from 4-8 letters. Secondly, the word/non-word by orthographic prediction error interaction was also reliable when the prediction included all words of all letter lengths from the English lexicon (see part b of this Figure) and the orthographic prediction error estimation was based on this length-unspecific prediction (estimate: 0.02; SE=0.007; t=3.349). (b) Letter-length unspecific prediction for English based on ~60,000 English words from the SUBTLEX database (Heuven et al., 2014).



Supplemental figure 2. Dutch lexical decision behavior and prediction using a proportional script. (a) Effect of the orthographic prediction error parameter, (b) number of pixels parameter and (c) showing the same model comparisons realized in Figure 3 for the data from the first Dutch lexicon project (DLP1; Keuleers, Diependaele, & Brysbaert, 2010; 4,305 five-letter stimuli; 39 participants) and the same effects and model comparisons for the second Dutch lexicon project (DLP2; Brysbaert, Stevens, Mandera, & Keuleers, 2016; 3,145 five-letter stimuli; 81 participants) are presented in (d,e,f). Before going into the details of the two studies one has to note that the patterns we have found in the data in relation to our parameters of interest do not replicated within these two Dutch studies and, in addition, do not replicate with the findings from German, English, and French shown in Figure 3. In general, this is difficult for the interpretations of the results. For the DLP1 pattern we found a significant interaction of the orthographic prediction error with word/non-words and no significant effect of number of pixels. The interaction pattern in contrast to the findings in other languages (Fig. 3a), however, was qualitatively different as it showed a negative orthographic prediction error effect for words and a positive effect for non-words. The pattern is exactly the inverse from all other languages. Still model comparisons highlighted that the orthographic prediction error was relevant for the model fit since the predictor increased the

model fit with no further increase of fit when the number of pixel parameter was included. None of these findings could be replicated in the DLP2 dataset, showing no significant fixed effects or interactions and no substantial changes in model fit relation to the null model. (g) Prediction image from a VOP implementation using five-letter words with a proportional Times New Roman script.



Supplemental figure 3. Detailed description of significant activation clusters for (a) the orthographic prediction error; (b) word/non-word effect; (c) interaction of word/non-word and the orthographic prediction error. On the left, the effect sizes from regression ERPs are presented as time courses for each sensor and time-point (color coding reflects scalp position). This part of the Figure reproduces Figure 4. The right column displays time courses with one line per channel, masked by significance using cluster statistics (see Methods for details; Maris & Oostenveld, 2007).

Supplemental table 1. Results from linear mixed model regression analysis (with the exception of the British data including multiple word lengths was estimated based on word aggregated data) for the behavioral lexical decision tasks (LDT) and handwriting analyses.

	E	SE	*t*
German LDT N°1: Orthographic prediction error based on word length specific prediction			
Intercept	6.49	0.023	288
Orthographic prediction error (oPE)	-0.03	0.004	6.5
Number of pixels (Npixel)	-0.007	0.004	1.8
Word/non-word (Lex)	0.33	0.009	33.1
Word frequency	-0.12	0.004	33.5
Error	-0.03	0.005	6.2
oPE X Lex	0.03	0.006	5.0
Npixel X Lex	0.000	0.006	0.1
German LDT N°1: Orthographic prediction error based on word length general prediction			
Intercept	6.48	0.023	288.3
Orthographic prediction error (oPE)	-0.03	0.004	6.3
Number of pixels (Npixel)	-0.01	0.004	1.7
Word/non-word (Lex)	0.33	0.010	33.2
Word frequency	-0.12	0.004	35.5
Error	-0.03	0.005	6.2
oPE X Lex	0.03	0.006	4.5
Npixel X Lex	-0.00	0.006	0.0
German LDT N°1: Orthographic prediction error based on word length specific prediction including orthographic Levenshtein distance and word frequency			
Intercept	6.66	0.023	237.1
Orthographic prediction error (oPE)	-0.02	0.004	4.3

Number of pixel (Npixel)	-0.00	0.004	0.2
Word/non-word (Lex)	0.29	0.011	27.0
Error	-0.03	0.005	6.2
Orthographic Levenshtein distance	-0.08	0.008	10.5
Word frequency	-0.12	0.004	35.5
oPE X Lex	0.03	0.006	5.2
Npixel X Lex	-0.00	0.005	0.6

German LDT N°2 including noise: 0%

Intercept	6.32	0.024	263.9
Orthographic prediction error (oPE)	-0.02	0.016	1.4
Number of pixels (Npixel)	-0.00	0.015	0.2
Word/non-word (Lex)	0.27	0.05	5.4
Word frequency	-0.07	0.02	4.9
Error	-0.07	0.010	6.8
oPE X Lex	0.05	0.02	2.3
Npixel X Lex	-0.02	0.021	1.2

German LDT N°2 including noise: 20%

Intercept	6.45	0.026	245.4
Orthographic prediction error (oPE)	-0.06	0.017	3.3
Number of pixels (Npixel)	-0.00	0.013	0.3
Word/non-word (Lex)	0.37	0.049	7.5
Word frequency	-0.14	0.02	6.1
Error	-0.14	0.010	5.4
oPE X Lex	0.04	0.022	1.6
Npixel X Lex	0.02	0.022	0.7

German LDT N°2 including noise: 40%

Intercept	6.84	0.042	162.9
Orthographic prediction error (oPE)	-0.02	0.021	1.0

Number of pixels (Npixel)	-0.08	0.018	4.1
Word/non-word (Lex)	0.14	0.049	2.8
Word frequency	-0.11	0.06	1.9
Error	-0.00	0.010	0.1
oPE X Lex	-0.00	0.028	0.1
Npixel X Lex	0.08	0.026	2.9

British LDT

Intercept	6.39	0.013	507.1
Orthographic prediction error (oPE)	-0.007	0.001	5.3
Number of pixels (Npixel)	0.008	0.001	6.7
Word/non-word (Lex)	0.12	0.003	46.2
Word frequency	-0.067	0.001	58.0
oPE X Lex	0.008	0.002	4.2
Npixel X Lex	-0.003	0.002	1.9

British LDT 4-8 Letters: Length specific prediction

Intercept	6.26	0.157	39.7
Orthographic prediction error (oPE)	-0.001	0.000	5.0
Number of letters (Nletters)	0.062	0.027	2.3
Word/non-word (Lex)	0.155	0.162	0.3
Error	0.043	0.165	0.8
oPE X Lex	-0.001	0.000	4.5
oPE X Nletters	-0.001	0.000	3.3
oPE X Error	-0.002	0.000	5.1
Nletters X Lex	-0.006	0.028	0.8
Nletters X Error	-0.245	0.172	1.4
Lex X Error	-0.036	0.028	1.3
oPE X Lex X Nletters	0.001	0.000	2.4
oPE X Lex X Error	0.002	0.000	5.0

oPE X Nletters X Error	0.001	0.000	3.2
Nletters X Lex X Error	0.003	0.030	0.1
oPE X Lex X Nletters X Error	-0.001	0.000	2.6

British LDT 4-8 Letters: Length general prediction

Intercept	5.25	0.421	12.5
Orthographic prediction error (oPE)	0.002	0.000	3.7
Number of letters (Nletters)	0.250	0.061	4.1
Word/non-word (Lex)	1.064	0.438	2.4
Error	1.264	0.443	2.9
oPE X Lex	-0.002	0.001	3.1
oPE X Nletters	-0.000	0.000	3.6
oPE X Error	-0.002	0.001	4.0
Nletters X Lex	-0.183	0.065	2.9
Nletters X Error	-0.002	0.001	4.0
Lex X Error	-1.426	0.467	3.1
oPE X Lex X Nletters	0.001	0.000	2.9
oPE X Lex X Error	0.002	0.001	3.6
oPE X Nletters X Error	0.001	0.000	4.0
Nletters X Lex X Error	0.228	0.068	3.5
oPE X Lex X Nletters X Error	-0.001	0.000	3.3

British LDT 4-8 Letters: Number of pixel

Intercept	6.590	0.157	42.0
Number of pixel (Npixel)	0.000	0.001	0.3
Number of letters (Nletters)	0.092	0.028	3.2
Word/non-word (Lex)	-0.124	0.162	0.8
Error	-0.309	0.165	1.9
Npixel X Lex	0.000	0.001	0.2

Npixel X Nletters	0.000	0.001	1.4
Npixel X Error	0.000	0.001	0.4
Nletters X Lex	-0.059	0.029	2.0
Nletters X Error	-0.090	0.030	3.0
Lex X Error	0.035	0.171	0.2
Npixel X Lex X Nletters	0.000	0.001	0.9
Npixel X Lex X Error	0.000	0.001	0.1
Npixel X Nletters X Error	0.000	0.001	1.2
Nletters X Lex X Error	0.069	0.031	2.2
Npixel X Lex X Nletters X Error	0.000	0.001	1.2

French LDT

Intercept	6.63	0.005	1,333
Orthographic prediction error (oPE)	-0.002	0.001	2.0
Number of pixels (Npixel)	0.002	0.001	1.3
Word/non-word (Lex)	-0.040	0.003	11.6
Word frequency	-0.042	0.001	34.1
oPE X Lex	0.005	0.002	2.0
Npixel X Lex	-0.007	0.002	3.0

Dutch LDT

Intercept	6.45	0.019	348.1
Orthographic prediction error (oPE)	0.005	0.002	3.2
Number of pixels (Npixel)	0.001	0.002	0.6
Word/non-word (Lex)	0.101	0.004	23.8
Word frequency	-0.061	0.002	36.9
oPE X Lex	-0.016	0.002	6.6
Npixel X Lex	0.002	0.002	1.0

Dutch LDT2

Intercept	6.35	0.016	391.1
Orthographic prediction error (oPE)	0.002	0.002	1.1
Number of pixels (Npixel)	-0.001	0.002	0.6
Word/non-word (Lex)	0.048	0.005	9.4
Word frequency	-0.023	0.001	26.9
oPE X Lex	-0.003	0.003	1.3
Npixel X Lex	0.003	0.003	0.5
Handwriting: Script based orthographic prediction error			
Intercept	1.465	0.010	154.3
Mean prediction strength	0.052	0.007	7.4
Number of pixels with a prediction	0.015	0.008	2.1
Letter case	0.039	0.012	3.2
Handwriting: Readability ratings			
Intercept	11.5	1.4	8.1
Mean prediction strength	-5.9	1.0	6.2

Note. E: Estimate; SE: Standard error; *t*: t-value. All *t*'s >2 are considered a significant effect.