SCIENTIFIC STUDIES OF READING
http://dx.doi.org/10.1080/10888438.2015.1104688



# Visual complexity in orthographic learning: Modeling learning across writing system variations

Li-Yun Chang[a], David C. Plaut[b], and Charles A. Perfetti[c]

[a]National Taiwan Normal University; [b]Carnegie Mellon University; [c]University of Pittsburgh

**ABSTRACT**

The visual complexity of orthographies varies across writing systems. Prior research has shown that complexity strongly influences the initial stage of reading development: the perceptual learning of grapheme forms. This study presents a computational simulation that examines the degree to which visual complexity leads to grapheme learning difficulty. We trained each of 131 identical neural networks to learn the structure of a different orthography and demonstrated a strong, positive association between network learning difficulty and multiple dimensions of grapheme complexity. We also tested the model's performance against grapheme complexity effects on behavioral same/different judgments. Although the model was broadly consistent with human performance in how processing difficulty depended on the complexity of the tested orthography, as well as its relationship to viewers' first-language orthography, discrepancies provided insight into important limitations of the model. We discuss how visual complexity can be a factor leading to reading difficulty across writing systems.



## Introduction

The study of reading development across writing systems has focused primarily on the principles governing mapping between graphemes and various linguistic units such as phonemes, syllables, and morphemes. Grapheme encoding itself has received relatively less attention. Graphemes are the basic units that distinguish among a language's written morphemes (e.g., single letters and letter combinations in alphabets/abjads, akshara in alphasyllabaries/syllabaries, and characters in morpho-syllabaries). Accurate, stable orthographic representations are required for associations to be reliably learned between visual forms and other aspects of language (Perfetti & Hart, 2002). Representations of the visual forms of graphemes are thus a critical beginning point of reading.

Visual complexity influences the development of orthographic representations, thus contributing to difficulty in learning to read. Orthographies with visually complex graphemes are also likely to contain a larger grapheme inventory, making learning that much more difficult (e.g., Nag, 2011; Nag & Snowling, 2012; Nag, Treiman, & Snowling, 2010). Here, we use computational modeling as a tool to examine the relationship between the visual complexity of graphemes and learning difficulty across writing systems.

The visual demands of grapheme processing, driven by the size of the grapheme inventory and the corresponding complexity of the graphemes, can pose a challenge to beginning learners. Empirical studies covering a wide range of orthographies have demonstrated that grapheme complexity is negatively correlated with grapheme identification efficiency (Liu, Chen, Liu, & Fu, 2012;

**CONTACT** David C. Plaut plaut@cmu.edu Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hssr.

Pelli, Burns, Farell, & Moore-Page, 2006). Learners are further challenged in mastering the complete inventory of graphemes, which varies in size across orthographies. In alphabetic orthographies (average number of graphemes: 20–30) such as Finnish, children master all graphemes after first grade (Seymour, Aro, & Erskine, 2003; White, Graves, & Slater, 1990); in alphasyllabic orthographies (average number of graphemes is 400) such as Kannada, children require 3 to 4 years of formal instruction to master all graphemes (Nag, 2007); in morphosyllabic orthographies (average number of graphemes is greater than 3,000) such as Chinese, children continue to learn novel characters after 6 years of formal education (Shu, Chen, Anderson, Wu, & Xuan, 2003). Thus, grapheme inventory size, ranging from quite small in "contained" orthographies to extremely large in "extensive" orthographies (Nag, 2007), has a strong influence on the pace of orthographic learning. The larger the inventory, the greater the degree of exposure over a longer period required to master the visual forms of the writing system.

Reading orthographies with large grapheme inventories and more complex graphemes may require stronger visual-spatial skills and may, in turn, strengthen such skills. Tan, Spinks, Eden, Perfetti, and Siok (2005) found that early progress in reading Chinese was linked more to copying skills than to phonemic awareness. Moreover, McBridge-Chang, Zhou, Cho, Aram, Levin, and Tolchinsky (2011) reported a link between the complexity of an orthography and children's visuospatial skills. Children learning to read traditional Chinese, which is written with highly complex graphemes (average 10 strokes per character; Huang & Hanley, 1995), outperformed children learning to read Spanish, which is written with relatively simple graphemes (average of 2.5 strokes per letter; Changizi & Shimojo, 2005), on a standardized visuospatial relationship task. A similar effect of grapheme complexity was found in beginning readers learning two orthographies. Abdelhadi, Ibrahim, and Eviatar (2011) compared visual vowel detection among Arabic–Hebrew bilingual children and reported that the same individuals had higher accuracy in Hebrew, a visually simple orthography, than in Arabic, a visually complex orthography. Furthermore, comparisons of performance of Chinese and English dyslexic readers on a similar visual spatial task found a parallel difference (Everatt, Jeffries, Elbeheri, Smythe, & Veii, 2006; McBride-Chang et al., 2013). These results establish the importance that visual complexity has on reading development. This impact has been explored within a variety of types of writing systems, including alphabets (Treiman & Kessler, 2011), alphasyllabaries (Nag, 2011), and morphosyllabaries (Shu et al., 2003). Our study aims to add a comparative perspective that can be applied across writing systems, using both a general characterization of visual complexity and a general model of grapheme learning.

Models of reading have tended to focus not on the visual forms used in writing but on the mapping of these forms to linguistic units: phonological transparency (e.g,, Harm & Seidenberg, 1999), the mapping between orthography and semantics (e.g., Harm & Seidenberg, 2004; Plaut, 1997; Plaut & Gonnerman, 2000) and reading words aloud (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Zevin & Seidenberg, 2006; see Frost, 2012; Seidenberg, 2011, for reviews). Several recent models have focused on alphabetic writing to examine the encoding of grapheme strings—the spatial coding model (Davis, 2010), the overlap model (Gomez, Ratcliff, & Perea, 2008), the Bayesian reader model (Norris, Kinoshita, & van Casteren, 2010), and the sequential encoding regulated by inputs to oscillating letter (SERIOL) model (Whitney, 2001). Models of reading Chinese (e.g., Perfetti, Liu, & Tan, 2005; Taft, 2006; Yang, McCandliss, Shu, & Zevin, 2009) have focused on grapheme forms but have required independent coding specific to Chinese, with no natural generalization to other orthographies. What is needed to capture the broader reality of human writing is an approach that does not stipulate the specific form of graphemic representation but can learn appropriate visual features through experience with any orthography. Such a universally applicable learning model would provide a basis for comparing the difficulty of graphemic form learning across writing systems and would provide the field with a more general picture of orthographic development in reading (see Frost, 2012). A general graphemic learning system could also provide a systematic assessment the complexity of any grapheme in the world's writing systems. Such a quantitative, objective tool would provide the basis for measurement of orthographies, enabling comparisons within and across writing systems and the testing of hypotheses about the perceptual processing of graphemes.

An important quantitative analysis of cross-orthographic grapheme complexity by Pelli et al. (2006) focused solely on *perimetric complexity*, defined as the ratio of (a) the square of perimeter length, measured both inside and outside curves, and (b) total ink area of a shape. Using this measure, Pelli and colleagues compared a number of orthographies, typefaces, and artificial symbol sets, concluding that perimetric complexity was a general predictor of learning efficiency. However, other visual characteristics of graphemes have been demonstrated to have perceptual salience in related reading studies, such as verticies (Lanthier, Risko, Stolz, & Besner, 2009), disconnected components of single graphemes (Winskel, 2009), and number of strokes (Tamaoka & Kiyama, 2013). Chang, Chen, and Perfetti (2015) developed a more comprehensive visual complexity measure by combining perimetric complexity with measurements involving these additional salient characteristics and applied the measure to quantify 131 orthographies (see Appendix A for details on the complexity measure, and Appendix B for detail on the orthographies).

The result was a clear trend across writing systems in how they handled variability in visual properties. In general, orthographies from a writing system with a larger grapheme inventory (e.g., alphasyllabary) tend to be more visually complex than those from a writing system with a smaller inventory (e.g., alphabet).

This visual complexity measure was applied to a cross-writing-system study that examined the effect of learning to read different orthographies on visual perceptual processing (Chang, 2014). The complexity measure was used to rank five orthographies, each representing a major writing system, from least to greatest in complexity: Hebrew (abjad), Russian (alphabet), Cree (syllabary), Telugu (alphasyllabary), and Chinese (morphosyllabary). With the addition of other orthographies to more fully represent variations in complexity—from visually simple to complex: Hebrew, English, Russian, Arabic, Hindi, Telugu, Japanese, and Chinese—eight orthographies were used in a perceptual judgment experiment involving 480 adult, age-matched native readers of these orthographies. In the task, participants saw two graphemes simultaneously and were instructed to quickly and accurately judge whether the graphemes were the same (identical) or different (see Appendix C for details about the cross-writing-system experiment, and Appendix D for example graphemes, varying in complexity, from each tested orthography). Participants' perceptual judgments were influenced both by grapheme complexity and by first-language (L1) background (see Figure 1). As grapheme complexity increased, the accuracy of perceptual judgments decreased, whereas individuals' L1 background interacted with grapheme complexity to produce patterns of performance that differed by L1. These results suggest that the visual complexity



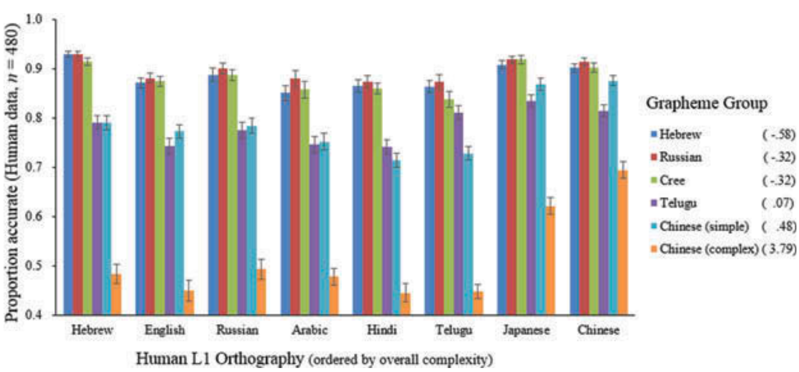


**Figure 1.** Proportion accurate of same-different judgments made for grapheme pairs drawn from different orthographies (grapheme group, plotted as different colored bars), made by participants with different first-language orthographies. Adapted from Chang (2014).

of both the perceived graphemes and L1 orthography affects perceptual processing of graphemes, and thus influences the fundamental stages of learning to read.

Although these findings on perceptual processing suggest that grapheme complexity plays a role in reading, there is a possible confounded factor: grapheme complexities covary with mapping principles between orthographic and linguistic units (Perfetti & Harris, 2013) in that more visually complex orthographies tend to map onto higher level linguistic units. Indeed, the need for complexity is driven by the size of the grapheme inventory, which in turn is driven by the size of linguistic units to which they map: phonemes, syllables, syllabic morphemes, in increasing order. In perceptual judgments, there is little reason to think that mapping principles are relevant, especially for orthographies not known to a participant. However, in the case of learning to read, the mapping principles are indeed very relevant, and individual mappings of graphemes to language units are the main object of early reading instruction. The learning of these mappings, however, is dependent on learning of the graphemes as visual forms.

To address the strictly visual aspect of grapheme learning, independent of mapping, we developed a computational model of grapheme learning that has no knowledge of linguistic units and, hence, is based solely on learning the visual forms of graphemes. It thus provides a means of exploring purely visually based, generalized grapheme learning. To be clear, we do not expect such a limited model to capture all aspects of human performance in this domain; rather, we view the model as a valuable means of evaluating and understanding both the strengths and limitations of the approach, with the goal of informing improved modeling efforts in the future.



## The model

We developed a model within the Parallel Distributed Processing (PDP) framework because of its abstract approximation of neural computation and its intrinsic ability to model learning (Plaut, 2005; Seidenberg, 2006) by changes in the model's output over time as a function of the input it receives. In a PDP model, processing takes the form of cooperative and competitive interactions over many simple processing units instead of activation of single units. Knowledge is encoded by weights on the connections among the units; learning involves iteratively adjusting the weights based on performance feedback. After learning, the model generalizes its knowledge to novel input, as determined by the similarity between the novel and learned representations. In short, PDP models instantiate learning as an incremental increase in knowledge. Such models have been used to simulate reading processes in English (e.g., Zevin & Seidenberg, 2006) and in Chinese (e.g., Yang et al., 2009), in skilled and less-skilled readers (e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989), and in normal and dyslexic readers (e.g., Harm & Seidenberg, 1999; Plaut, 1999; Woollams, Lambon Ralph, Plaut, & Patterson, 2007). Recent work (Di Bono & Zorzi, 2013; Hannagan, Ziegler, Dufau, Fagot, & Grainger, 2014; Zorzi, Testolin, & Stoianov, 2014) has applied specialized multilayer networks—so-called deep learning or convolutional neural networks (Hinton, Osindero, & Teh, 2006; see LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015, for reviews)—to learning orthographic representations of letter strings within a single orthography.

Our aim was to explore the extent to which a general computational model would capture the effect of visual complexity of graphemes on orthographic learning. To that end, we applied the same basic functional architecture to learning each of 131 orthographies to test the effect of complexity on learning outcomes. We also modeled Chang's (2014) perceptual judgment experiment to provide an account of how the visual complexities of graphemes and the L1 orthography to which they belong both affect perceptual processing. To the best of our knowledge, this is the first attempt to model grapheme learning over a large number of orthographies and to examine the extent to which visual complexity affects learning to read across writing systems.

## Method

Additional details of the modeling methodology that are not included here are given in Appendix E. All network and example files and simulation software is available for download at http://www.cnbc.cmu.edu/~plaut/ChangPlautPerfetti-SSR.

### *Model architecture*

The model is a specific form of three-layer neural network known as an *autoencoder*. A standard autoencoder learns to copy patterns of activity over a group of input units onto an identically sized group of output units via a smaller number of intermediate or "hidden" units. Because there are fewer hidden units than input (or output) units, the network must learn to re-represent the inputs in a more concise form. In this way, the hidden representations come to emphasize the underlying structure shared by the ensemble of inputs at the expense of more idiosyncratic aspects of only one or a few patterns (Hinton & Salakhutdinov, 2006).

The architecture of the specific network used in the current work is depicted in Figure 2. The input patterns are images of graphemes over the 38 × 38 array of units at the bottom of the figure. Note that, because the input and output groups have exactly the same structure, only a single group of units is shown. The hidden layer is divided into four groups of units (shown at the top of Figure 2) that differ in number of units and in the sizes of their "receptive fields" (RFs). In particular, each hidden unit receives input only from a restricted circular region of the input and projects to the corresponding circular region of the output (these are depicted in red for four representative units). To allow the network to learn to be sensitive to features of varying scales and positions, different groups of units had different RF sizes, with centers spaced evenly across the input (and output) arrays: a 19 × 19 group with a RF diameter of five units and centers spaced every two units horizontally and vertically; a 12 × 12 group with RF diameter = 7 and spacing = 3; a 9 × 9 group with RF diameter = 11 and spacing = 4; and a 7 × 7 group with RF diameter = 15 and spacing = 5. Including "bias" connections (which determine the activation of units in the absence of other inputs), the network had a total of 83,607 connections. As a point of comparison, if all 635 hidden units were fully connected to both the input and output, the network would have required 1,835,959 connections. Using topographically restricted connectivity not only drastically reduces the required number of connections, and is broadly compatible with patterns of connectivity in visual cortex, but also encourages the network to discover largely local features of varying scales. Note that, because units in each hidden group are free to develop distinct receptive and projective fields, the network is more flexible than a standard convolutional neural network (in which each such unit is constrained to have identical incoming and outgoing weights). On the other hand, such networks typically employ multiple hidden layers between input and outputs, whereas our model has only a single hidden layer (with a range of receptive field sizes).

### *Stimuli*

Two sets of stimuli were adapted from Chang (2014). The first set served as a training set to simulate L1 orthographic learning; the second set served as a testing set to simulate human behavior in the same-different judgment.

The training patterns were composed of grapheme images from 131 orthographies across five writing systems (alphabetic = 60, alphasyllabary = 41, abjad = 16, syllabary = 11, morphosyllabary = 3); a total of 21,821 graphemes were used. To generate these images, the programming language Processing (Reas & Fry, 2010) was used to construct a simple image of each individual grapheme. Graphemes were presented in Arial font, in white against a 38 × 38-pixel black background. Of the selected orthographies, 25% did not have support for Arial font; for these, an alternative font similar to Arial was adopted (see Chang et al., 2015, for the fonts). These images were converted to 8-bit

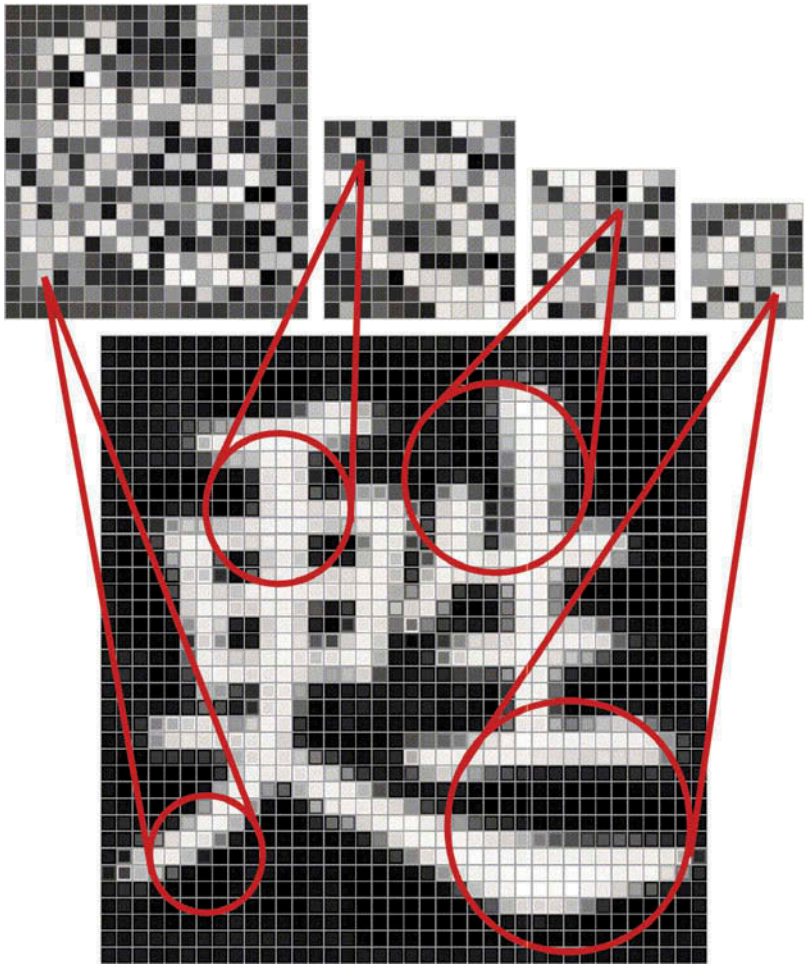

**Figure 2.** The architecture of the model used in the simulation. *Note*. Each small square corresponds to a unit. Input is presented as activity values (shown in grayscale, with black = 0.0 and white = 1.0) over the 38 × 38 array at the bottom; four groups of hidden units, varying in number and in receptive field size and spacing, are shown at the top. Input-to-hidden and hidden-to-output connections were restricted to topographically constrained circular "receptive fields"—the red lines depict the scale of these receptive fields for four representative hidden units (no actual connections are shown). The output units have exactly the same 38 × 38 form as the input units and are not depicted separately; rather, their activations (for an example complex Chinese character after training) are shown in the central region of each input unit, with the actual input value shown in the surrounding ring. Thus, units for which the center and surround match are fully accurate in their reconstructed activations.

integer values, then inverted and normalized to values between 0.0 and 1.0, forming training patterns to be passed to the network just described.

Testing patterns were generated in the same manner as the training patterns. Identical to Chang's (2014) experiment, grapheme pairs were selected within an orthography (i.e., Hebrew, Russian, Cree, Telugu, and Chinese), matched to either uppercase or lowercase and, where applicable, vowel or consonant (for Telugu vowels, primary forms were matched with primary forms and vowel diacritics were matched with vowel diacritics); all graphemes in each orthography were exhaustively used (except for Chinese). Given that thousands of graphemes with highly variable complexity exist in the Chinese orthography, two groups of graphemes with contrasting complexity (simple or complex) were formed from the overall orthography, and pairs were selected within each group. Each "simple" character was a radical, the functional "building block" in Chinese orthography (Shen & Ke, 2007), composed of a small number of strokes (average =

4.52); "complex" characters were those containing multiple radicals, composed of a large number of strokes (average = 13.21). Note that these characters shared the same forms between the traditional and simplified Chinese visual orthographies. In summary, the testing patterns included six grapheme groups, each with a similar number of grapheme pairs and a varying overall complexity value (standardized score) from Chang et al.'s (2015) measurement system: Hebrew (256; –.58), Russian (264; –.32), Cree (320; –.32), Telugu (280; .07), Chinese simple characters (160; .48), and Chinese complex characters (160; 3.79).

### Training

In the real world, successful orthographic learning critically involves successful visual representation and recognition of graphemes. In the model, successful learning occurs when hidden units develop an internal representation of each input image that enables it to be reconstructed over the output layer with minimal difference between the inputs and outputs, that is, minimal reconstruction error. To reduce reconstruction error, the back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986) was used in the present model with online learning, a learning rate of 0.01 and momentum of 0.8. To simulate orthographic learning across writing systems, 10 versions of the network (varying only in initial random weights) were trained on the grapheme patterns from each of 131 orthographies, with number of patterns corresponding to their grapheme inventory size, resulting in 131 trained networks. For each, training was halted when the average reconstruction error across the entire set of graphemes in that orthography fell below a fixed criterion of 10.0.

The number of learning epochs required in reaching the average error per grapheme of 10.0, averaged over the 10 versions of the network, was taken as the primary measure of the difficulty of learning a given orthography. The use of an error measure that is averaged (rather than summed) over graphemes is important because it allows us to reasonably compare orthographies that differ widely in number of graphemes.

### Testing

To model how individuals with different L1 experiences approach graphemes with various complexities, we first selected eight trained sets of networks (average error < 10) to represent skilled L1 readers as in Chang's (2014) experiment. Next, we presented these networks (i.e., Hebrew, English, Russian, Arabic, Hindi, Telugu, Japanese, and Chinese) with testing patterns consisting of pairs of both identical and differing graphemes, taken from six grapheme groups (i.e., Hebrew, Russian, Cree, Telugu, simple Chinese characters, and complex Chinese characters). Each grapheme in a pair was presented separately to the network, and the activation values over all 635 hidden units were recorded. We assume that same-difference judgments are made on the basis of the similarity of these representations. To measure this similarity, we repeatedly added noise to each element of the two hidden patterns, computed their correlation, and then averaged the results (n.b., noise was added to both "same" and "different" trials because the hidden representations of the "same" trials were identical). We used these averaged correlations between hidden patterns within each grapheme group to characterize each model's performance, as an instantiation of human performance. Higher correlations indicated worse performance on "different" trials (e.g., longer reaction time or lower accuracy) because the two given graphemes were more similar and thus more difficult to discriminate; lower correlations indicated better performance. Finally, to approximate the accuracy data in the behavioral experiment, these correlations were inverted by subtracting them from 1 to create a measure termed *representation dissimilarity*, which we compare against human accuracy data.

## Results

### *Effects of grapheme complexity on learning time*

In training networks to learn internal representations of graphemes, we were particularly interested in the relationship between grapheme complexity and learning performance. As an initial analysis, we correlated the number of learning epochs (to reach average error < 10) from the 131 networks (10 per orthography) with the overall complexity measures of the 131 corresponding orthographies from Chang et al. (2015). This resulted in a significant correlation of .66 ($p < .001$, two-tailed), indicating that grapheme complexity and learning difficulty were strongly, positively associated.

As discussed earlier, however, grapheme complexity is correlated with the size of the grapheme inventory ($r = .71$, $p < .001$). To attempt to tease these factors apart, we carried out a partial regression of both the complexity measure and inventory size against training epochs. Together the factors produced a correlation of .80 with training epoch, with both complexity ($\beta = .20$), $t(128) = 2.67$, $p < .001$, and inventory size ($\beta = .64$), $t(128) = 8.42$, $p < .001$, accounting for unique variance. Thus, and not surprising, orthographies with more graphemes require greater training time to reach criterion (even when based on average grapheme reconstruction error), although visual complexity makes an independent contribution above and beyond this effect.

In addition, given the relatively unstructured nature of the model, it is informative to consider whether the strong relationship between training times and grapheme complexity is driven by particular dimensions of the complexity measure (see Appendix A). In particular, it might be that the network is particularly sensitive to parametric complexity (PC) relative to the other dimensions (numbers of simple features [SF], connected points [CP], and disconnected components [DC]) as it is more easily computed from pixel-based input without counting discrete entities. As it turns out, however, a partial regression of these dimensions against training time reveals that each accounts for unique variance, and PC is actually the weakest predictor: PC ($\beta = .24$), $t(126) = 2.61$, $p < .05$; SF ($\beta = -1.30$), $t(126) = 5.82$, $p < .001$; CP ($\beta = 1.29$), $t(126) = 7.02$, $p < .001$; DC ($\beta = .62$), $t(126) = 6.33$, $p < .001$. Moreover, the weighted combination of dimensions has a higher correlation with training epochs ($r = .77$, $p < .001$) than does the simple averaging used in our complexity measure ($r = .66$, $p < .001$).

Finally, it is worth considering whether the same results might be obtained based solely on a simple measure of input similarity without any contribution of learned internal representations. To evaluate this possibility, we computed the mean similarity (Pearson correlation coefficient) among all nonidentical pairs of graphemes within each orthography and compared this measure to the grapheme complexity measure and to network training times. Mean pairwise similarity was, in fact, moderately (negatively) correlated with training time ($r = -.22$), $t(129) = 2.61$, $p < .05$. However, whereas training time was strongly related to graphemic complexity ($r = .66$), pairwise similarity was unrelated to graphemic complexity ($r = .04$), $t(129) = 0.54$, *ns*. Similar results hold for pixelwise entropy (thresholding values at 0.5), which is another surface-level characterization of information content: Entropy was moderately correlated with training time ($r = .28$), $t(129) = 3.31$, $p < .01$, but not with graphemic complexity ($r = -.11$), $t(129) = 1.27$, $p = .20$. Thus, the network's learned representations are making an important contribution to performance beyond mere sensitivity to pixelwise input similarity or entropy.

### *Effects of grapheme complexity and L1 background on perceptual judgments*

Turning to a consideration of the findings from the same-different judgment task (Chang, 2014), an initial question was whether, for the eight L1 networks comparing graphemes from six grapheme groups, L1 background affects perceptual variability across grapheme complexity levels. To address this question, an 8 × 6 (L1 Background × Grapheme Complexity) analysis of variance was conducted with network accuracy (representation dissimilarity) as the dependent measure. The main effect of

**Table 1.** Means and standard deviations (in parentheses) of network accuracy in examining whether L1 orthographic complexity differentially affects perceptual variability across grapheme complexity levels.

| Network (Ordered by trained orthographies from simple to complex) | Grapheme group (Ranked complexity from simple to complex) 1 Hebrew | 2 Russian | 3 Cree | 4 Telugu | 5 Simple Chinese | 6 Complex Chinese | $F$ | $\eta_p^2$ | Pairwise comparison (With bonferroni adjustments) |
|---|---|---|---|---|---|---|---|---|---|
| Hebrew (–.58) | 0.44 (0.003) | 0.45 (0.003) | 0.51 (0.005) | 0.45 (0.014) | 0.38 (0.005) | 0.34 (0.006) | 1236.03** | .993 | 3 > 1,2,4 > 5 > 6 |
| English (–.50) | 0.44 (0.003) | 0.45 (0.002) | 0.53 (0.004) | 0.44 (0.007) | 0.39 (0.004) | 0.36 (0.004) | 2907.04** | .997 | 3 > 2 > 1,4 > 5 > 6 |
| Russian (–.32) | 0.43 (0.003) | 0.46 (0.002) | 0.50 (0.003) | 0.42 (0.007) | 0.38 (0.004) | 0.35 (0.003) | 2493.22** | .996 | 3 > 2 > 1,4 > 5 > 6 |
| Arabic (–.26) | 0.41 (0.004) | 0.42 (0.005) | 0.51 (0.006) | 0.43 (0.008) | 0.36 (0.004) | 0.30 (0.004) | 4050.83** | .998 | 3 > 2,1,4 > 5 > 6 |
| Hindi (–.02) | 0.41 (0.003) | 0.44 (0.002) | 0.52 (0.004) | 0.44 (0.008) | 0.38 (0.004) | 0.31 (0.002) | 4424.54** | .998 | 3 > 2,4 > 1 > 5 > 6 |
| Telugu (.07) | 0.41 (0.004) | 0.43 (0.004) | 0.53 (0.005) | 0.49 (0.009) | 0.38 (0.004) | 0.31 (0.004) | 3717.38** | .998 | 3 > 4 > 2 > 1 > 5 > 6 |
| Japanese (1.62) | 0.45 (0.003) | 0.49 (0.002) | 0.52 (0.003) | 0.43 (0.006) | 0.41 (0.002) | 0.41 (0.001) | 2426.23** | .996 | 3 > 2 > 1 > 4 > 5 > 6 |
| Chinese (3.79) | 0.45 (0.010) | 0.48 (0.011) | 0.51 (0.008) | 0.44 (0.013) | 0.42 (0.013) | 0.40 (0.021) | 199.952** | .913 | 3 > 2 > 1 > 4,5 > 6 |

*Note*. For the pairwise comparison with Bonferroni adjustments, all *ps* < .001.
**$p < .001$.

L1 background was significant, $F(7, 82) = 158.89$, $p < .001$, $\eta_p^2 = .931$; the main effect of grapheme complexity was also significant, $F(7, 410) = 5696.62$, $p < .001$, $\eta_p^2 = .986$. There was a significant interaction between L1 background and complexity, $F(35, 410) = 94.28$, $p < .001$, $\eta_p^2 = .889$. To better understand this interaction, we carried out pairwise comparisons using Bonferroni adjustments to control for the overall Type I error. Table 1 provides a summary of the comparisons along with the means and standard deviations of network accuracy.

Results revealed an accuracy gradient showing that increasing grapheme complexity leads to decreasing accuracy for all networks. Specifically, complex Chinese characters were the most difficult to process, followed by simple Chinese characters, and then three orthographies (i.e., Telugu, Russian, and Hebrew) with varying processing difficulty depending on the networks' trained orthographies; note that Cree graphemes were the least difficult to process, regardless of trained L1. Figure 3 illustrates that network accuracy for each tested grapheme group is a function of trained L1 orthography.

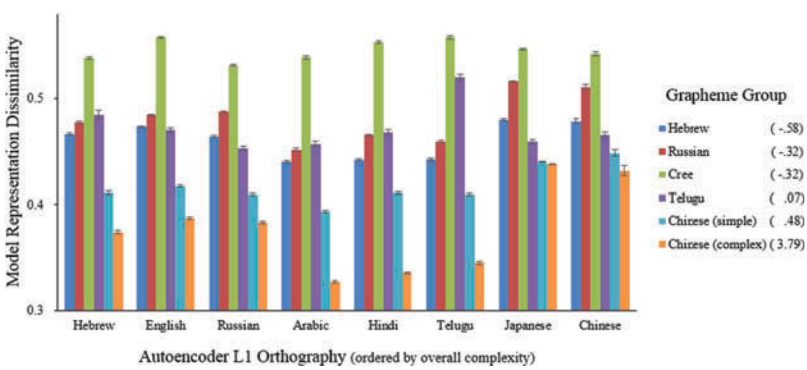


**Figure 3.** Modeling results of same-different judgments, in terms of representation dissimilarity, for grapheme pairs drawn from different orthographies (grapheme group, plotted as different colored bars), made by networks trained with different orthographies.

To answer the question of whether L1 background affects perceptual processing, we first examined the patterns from networks tested on graphemes from their original training orthography (i.e., Hebrew, Russian, Telugu, and Chinese networks). For the Hebrew networks, the Hebrew graphemes were just as difficult as Russian and Telugu graphemes and less difficult than simple and complex Chinese characters. For the Russian networks, however, the Russian graphemes were less difficult than Hebrew and Telugu graphemes, whereas the decreasing accuracy gradient was maintained for other grapheme groups. Similar to this L1 advantage of the Russian network, the Telugu networks showed less difficulty with Telugu graphemes than with other grapheme groups. Notably, although the Chinese networks were tested within their L1, their L1 advantage was not obvious. Complex Chinese characters remained the most difficult, whereas simple Chinese characters became as difficult as Telugu graphemes but not less difficult than Hebrew, Russian, and Cree graphemes (ordered in decreasing difficulty). Overall, these results suggested that, although familiarity played a role in networks' difficulty when processing L1, comparable to the effects of familiarity observed in Chang's (2014) behavioral results, grapheme complexity seemed generally to outweigh these effects in accounting for variability.

Next, we examined the patterns for networks tested on non-L1 graphemes (e.g., English, Arabic, Hindi, and Japanese networks). Across the four sets of networks (10 per orthography), a decreasing accuracy gradient was found for complex Chinese characters, simple Chinese characters, and Cree graphemes. However, processing difficulty with Telugu, Russian, and Hebrew graphemes varied across networks. For the English networks, Telugu and Hebrew graphemes were more difficult than Cree graphemes; for the Arabic networks, these three grapheme groups were equally difficult; for the Hindi networks, Hebrew graphemes were more difficult than Telugu and Russian graphemes; finally, for the Japanese networks, Telugu graphemes were more difficult than Hebrew graphemes and Russian graphemes. Although the variability in accuracy patterns from processing Telugu, Russian, and Hebrew graphemes needs further investigation, it is important to note that the general pattern holds across all six grapheme groups—increasing complexity leads to increased processing difficulty.

Collectively, results from the eight sets of networks establish a complexity effect—processing accuracy decreased as grapheme complexity increased, whereas complexity of L1 orthography interacted with tested grapheme complexity to produce patterns that differed by trained L1. For purposes of comparison, it is worth considering whether simple pairwise similarity among graphemes would exhibit the same pattern of results (apart from effects of L1 background, of course). To determine this, we calculate the pixelwise correlations among pairs of graphemes that occurred in "different" trials in the experiment and then subtracted this from 1.0 to provide a measure of the relative ease of correctly discriminating the graphemes. Across L1 backgrounds, the empirical data pattern into three groups of roughly equivalent difficulty: Hebrew, Russian, and Cree are easiest, following by Telugu and simple Chinese characters, with complex Chinese characters being much more difficult. The pairwise similarity measure, by contrast, patterns completely differently: Cree is easiest (0.964), following by Hebrew (0.946), complex Chinese (0.943) and simple Chinese (0.941), then Russian (0.921), with Telugu (0.828) being by far the most difficult. Thus, the pattern of complexity effects exhibited by both participants and the model cannot be explained by simple pixelwise similarity among graphemes.

At a general level, our modeling results are consistent with Chang's (2014) findings. However, close comparison of the empirical results (Figure 1) and modeling results (Figure 3) reveals some clear discrepancies as well. The most obvious is that the networks find Cree much easier to discriminate than any of the other orthographies, whereas for participants, it is only among the easiest (along with Hebrew and Russian). One possible explanation for this is that Cree graphemes are highly geometric shapes that often differ only by rotation or reflection (see Figure 4). Participants find these somewhat more difficult to distinguish because, based on extensive visual experience, their visual systems have learned to ignore transformations that preserve shape: translation, rotation, scaling, and reflection. By contrast, the network's representations did not develop under the pressure to recognize shapes across these transformations but developed only to reconstruct them accurately

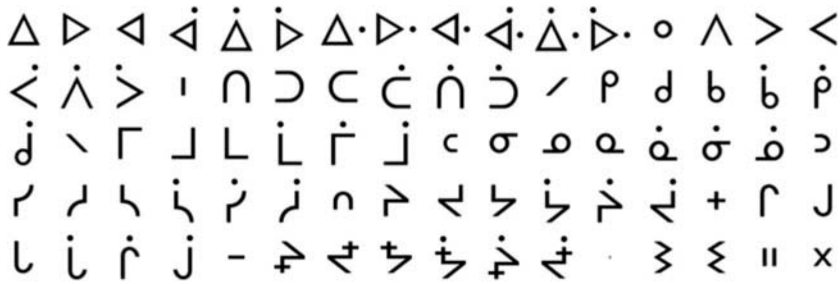

**Figure 4.** Cree graphemes.

on a pixelwise basis. As a result, the network outperforms participants because its representations for graphemes that differ only in these ways are overly distinct. A similar explanation likely applies to the network's relatively elevated performance (compared to participants) on Telugu graphemes, as they also have a high degree of similarity under rotation or reflection.

The other main discrepancies between the model and participants concern effects of familiarity, that is, being tested on the same orthography on which the model was trained. In cases such as Russian and Telugu, the network shows a somewhat greater benefit from familiarity than participants, again perhaps because it is overly sensitive to pixelwise similarity. On the other hand, the network shows less benefit from familiarity than participants when tested on the simpler Chinese characters. This may reflect inadequacies in our training procedure—in particular, failing to include relative grapheme frequency—but it may also reflect participants' partial reliance on higher-level (semantic) representations that the model lacks.

## Discussion

We developed a neural network model intended to capture basic visual characteristics of grapheme learning independent of specific orthographies, and then tested the model's learning of graphemes across a large number of orthographies. Using an autoencoder to implement a distributed-coding scheme, the model simulated grapheme learning in each of 131 orthographies, showing that grapheme complexity is positively, strongly associated with learning difficulty. Moreover, we simulated skilled learners from eight L1 orthographies performing same-different judgments on grapheme stimuli from six grapheme groups. Apart from some specific discrepancies involving Cree and Telugu (discussed next), the model captured the rough ordering of difficulty across orthographies, with Hebrew and Russian as the easiest (and least sensitive to L1 orthography), complex Chinese characters as the most difficult (and most sensitive to L1 orthography), and with simple Chinese characters falling somewhere in between. These simulations confirmed that perceptual performance is a function of complexity of grapheme stimuli and learners' L1 orthography.

The model's most important quality is its distributed-coding scheme for orthographic representations. Unlike the slot-coding scheme used commonly in prior modeling work (e.g., interactive activation model, McClelland & Rumelhart, 1981; dual route cascaded model, Coltheart et al., 2001), graphemes in this model were not individually represented with specified units. Each grapheme is represented by a particular pattern of activity over many units. Thus, our distributed coding scheme was designed to learn to detect features that make up graphemes in general. The fact that models trained on one orthography were accurate at representing and discriminating between graphemes from other orthographies suggests that the developed coding schemes have a fair degree of generality. Another important property of the model is the use of hidden units with varying receptive (and projective) field sizes; this design assured the model would not be biased to learning structure at any particular spatial scale (which might vary across orthographies). These capacities are critical because they allow for practical and precise cross-orthographic comparisons, which are important in capturing both the general and orthography-specific properties that combine to affect orthographic learning across the world's writing systems. Essentially, the present model serves as a universal grapheme learning device.

The model results are informative about the factors contributing to the visual phases of orthographic learning and to the perception of grapheme forms. First, simulating L1 learning across 131 orthographies showed that the number of learning epochs to reach mastery has a strong, positive correlation with grapheme complexity, as quantified by Chang et al. (2015) measurement system, and that this effect is not simply a matter of larger grapheme inventory or increased pixelwise similarities. This finding is in line with reports from vision research that the perceptual load of letters hinders individuals' recognition efficiency given individuals' limited visual-processing capacity (Pelli et al., 2006; Vogel, Woodman, & Luck, 2001; Xu & Chun, 2006). This interpretation is also consistent with several reading studies suggesting that perceptual load of visual orthography may be a source of processing difficulty (e.g., Nag, Snowling, Quinlan, & Hulme, 2014, for Kannada; Rao, Vaid, Srinivasan, & Chen, 2011, for Urdu). At the same time, the fact that grapheme inventory does make an independent contribution to learning times mirrors the prominent role-played by this factor in the pace of orthographic learning across writing systems (e.g., contained and extensive orthographies; Nag, 2007).

Second, the model provided a closer causal link from visual complexity of both graphemes and readers' L1 orthography to processing difficulty. Consistencies found when comparing the modeling results (Figure 3) with behavioral data in Chang (2014; Figure 1) suggest that processing difficulty increases as grapheme complexity increases while the specific pattern of difficulty differs by L1 orthography. Because our model focused on learning orthographic representation without mapping to phonology or semantics, the resulting patterns can be attributed to the complexity encoded in the representation. It is important that the model's orthography-focused design may also help to clarify the interpretation of the results of previous cross-orthography studies. Although these studies suggested that grapheme complexity plays a role in reading development (e.g., Abdelhadi et al., 2011; McBride-Chang et al., 2011), the confounding of complexity with level of linguistic mapping made the suggestion uncertain. Our simulations suggest that when linguistic mapping is out of the picture, there remains a strong effect of visual complexity.

In the broader context of learning to read, the effects of visual complexity cannot be fully established based solely on observations at the level of graphemes or orthography. Reading is fundamentally a process of associating graphemic forms with phonology and semantics, but the "stimulus encoding" phase is an important part of this process. In testing the hypothesis that complexity leads to reading difficulty and generalizing the results to various writing systems, a first step is to demonstrate that more visually complex orthographies are processed less reliably and efficiently, adding pressure when learning to read. This is what the current work has shown.

Notwithstanding the usefulness of the model in capturing general effects of graphemic complexity and L1 background, it must also be acknowledged that there were a number of substantial discrepancies between model and human performance when compared in detail. Graphemes from certain orthographies—Cree and Telugu—were too easy for the network to discriminate, and the network showed some effects of familiarity that differed from those of participants. Of importance, these discrepancies can be understood in terms of limitations of the current modeling approach. The model lacks the range of visual experience and task demands that, in participants, lead to substantial invariance to changes in translation, rotation, scaling, and reflection—and, hence, poorer performance when graphemes differ only in these respects. Indeed, a fully adequate account of perceptual invariances of these sorts will almost certainly require a much deeper and more structured network architecture (see LeCun et al., 2015; Schmidhuber, 2015). The network's inadequate treatment of familiarity may also reflect undue sensitivity to pixelwise similarities, as well as a lack of interactions with higher level representations. The identification of these limitations thus provides important information in the development of more comprehensive models of grapheme acquisition in particular, and visual processing in reading more generally.

It is also worth pointing out that, to fully account for graphemic learning in reading acquisition, the model would need to be extended to apply to strings of graphemes. Here we envision an approach similar to that used by Plaut (1999), in which the visual system initially fixates every grapheme (as in the current model) but then, as it develops competence in representing and processing such graphemes, begins to extend its processing—some might call it attention—to adjacent but slightly more peripheral graphemes as well. In this way the system gradually builds up the capability of representing and processing longer and longer strings of graphemes as it gains reading experience.

In conclusion, we have presented a novel approach to understanding the visual component of learning to read across writing systems by implementing a PDP learning model that represents orthographic knowledge in a distributed fashion. The model captures broad effects of graphemic complexity and L1 background on perceptual processing, and its limitations provide clear directions for future work. The broader contribution of this study is threefold. Theoretically, it shines a light on the visual processes that are important but often ignored in reading. Methodologically, we demonstrate the value of a distributed-coding scheme that can accommodate graphemes from any orthography, encouraging examination into orthographic learning across writing systems. Practically, we explain how visual complexity adds pressure to perceptual processing during reading, bringing to light a prominent risk factor for reading difficulty.



## Funding

This work was funded by a grant from the National Science Foundation (NSF) award SBE-0836012 through LearnLab, a grant from the NSF award BCS-1354350 to D. Plaut and M. Behrmann, and a grant from the Ministry of Education and the Ministry of Science and Technology in Taiwan (MOST 104-2911-I-003-301). This research was carried out while the first author was a Ph.D. student in the Department of Psychology at the University of Pittsburgh. The first author is indebted to Joseph Stafura and Xiaoping Fang for their helpful discussions during preparing the first version of the manuscript. The authors thank the anonymous reviewers and the editors for their comments and suggestions.

## References

Abdelhadi, S., Ibrahim, R., & Eviatar, Z. (2011). Perceptual load in the reading of Arabic: Effects of orthographic visual complexity on detection. *Writing Systems Research*, *3*, 117–127. doi:10.1093/wsr/wsr014

Ager, S. (1998). *Omniglot: The online encyclopedia of writing systems and languages*. Retrieved from http://www.omniglot.com

Chang, L. Y. (2014). *Visual orthographic variation and learning to read across writing system* (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.

Chang, L. Y., Chen, Y. C., & Perfetti, C. A. (2015). *Visual complexity measure and orthographic learning across writing systems*. Manuscript submitted for publication.

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*, 267–275. doi:10.1098/rspb.2004.2942

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*, E117–139.

Chen, H. C., Chang, L. Y., Chiou, Y. S., Sung, Y. T., & Chang, K. E. (2011). Construction of Chinese orthographic database for Chinese character instruction. *Bulletin of Educational Psychology*, *43*, 269–290.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608. doi:10.1037/0033-295X.100.4.589

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*, 204–256. doi:10.1037/0033-295X.108.1.204

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*, 713–758. doi:10.1037/a0019738

Di Bono, M. G., & Zorzi, M. (2013). Deep generative learning of location-invariant visual word recognition. *Frontiers in Psychology*, *4*, 1–10. doi:10.3389/fpsyg.2013.00635

Everatt, J., Jeffries, S., Elbeheri, G., Smythe, I., & Veii, K. (2006). Cross language learning disabilities and verbal versus spatial memory. *Cognitive Processing*, *7*, 32. doi:10.1007/s10339-006-0054-x

Frost, R. (2012). Towards a universal model of reading. *The Behavioral and Brain Sciences*, *35*, 263–279. doi:10.1017/S0140525X11001841

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Meredith.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review*, *115*, 577–600. doi:10.1037/a0012667

Hannagan, T., Ziegler, J. C., Dufau, S., Fagot, J., & Grainger, J. (2014). Deep learning of orthographic representations in baboons. *PLoS ONE*, *9*(1), e84843. doi:10.1371/journal.pone.0084843

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528. doi:10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720. doi:10.1037/0033-295X.111.3.662

Hinton, G. E., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*, 1527–1554. doi:10.1162/neco.2006.18.7.1527

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*, 504–507. doi:10.1126/science.1127647

Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition*, *54*, 73–98. doi:10.1016/0010-0277(94)00641-W

Lanthier, S. N., Risko, E. F., Stolz, J. A., & Besner, D. (2009). Not all visual features are created equal: Early processing in letter and word recognition. *Psychonomic Bulletin & Review*, *16*, 67–73. doi:10.3758/PBR.16.1.67

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. doi:10.1038/nature14539

Liu, T., Chen, W., Liu, C. H., & Fu, X. (2012). Benefits and costs of uniqueness in multiple object tracking: The role of object complexity. *Vision Research*, *66*, 31–38. doi:10.1016/j.visres.2012.06.009

McBride-Chang, C., Shu, H., Chan, W., Wong, T., Wong, A. M., Zhang, Y., … Chan, P. (2013). Poor readers of Chinese and English: Overlap, stability, and longitudinal correlates. *Scientific Studies of Reading*, *17*, 57–70. doi:10.1080/10888438.2012.689787

McBride-Chang, C., Zhou, Y., Cho, J.-R., Aram, D., Levin, I., & Tolchinsky, L. (2011). Visual spatial skill: A consequence of learning to read? *Journal of Experimental Child Psychology*, *109*, 256–262. doi:10.1016/j.jecp.2010.12.003

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading*, *30*, 7–22. doi:10.1111/jrir.2007.30.issue-1

Nag, S. (2011). The Akshara languages: What do they tell us about children's literacy learning? In R. K. Mishra & N. Srinivasan (Eds.), *Language-cognition interface: State of the art* (pp. 291–310). Munchen, Germany: Lincom.

Nag, S., & Snowling, M. J. (2012). Reading in an alphasyllabary: Implications for a language-universal theory of learning to read. *Scientific Studies of Reading*, *16*, 404–423. doi:10.1080/10888438.2011.576352

Nag, S., Snowling, M., Quinlan, P., & Hulme, C. (2014). Child and symbol factors in learning to read a visually complex writing systems. *Scientific Studies of Reading*, *18*, 309–324. doi:10.1080/10888438.2014.892489

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research*, *2*, 41–52. doi:10.1093/wsr/wsq001

Norris, D., Kinoshita, S., & van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language*, *62*, 254–271. doi:10.1016/j.jml.2009.11.002

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*, 4646–4674. doi:10.1016/j.visres.2006.04.023

Perfetti, C. A., & Harris, L. N. (2013). Universal reading processes are modulated by language and writing system. *Language Learning and Development*, *9*, 296–316. doi:10.1080/15475441.2013.813828

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam, the Netherlands: John Benjamins.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review*, *112*, 43–59. doi:10.1037/0033-295X.112.1.43

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, *12*, 765–806. doi:10.1080/016909697386682

Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, *23*, 543–568. doi:10.1207/s15516709cog2304_7

Plaut, D. C. (2005). Connectionist approaches to reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 24–38). Oxford, UK: Blackwell.

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*, 445–485. doi:10.1080/01690960050119661

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115. doi:10.1037/0033-295X.103.1.56

Rao, C., Vaid, J., Srinivasan, N., & Chen, H.-C. (2011). Orthographic characteristics speed Hindi word naming but slow Urdu naming: Evidence from Hindi/Urdu biliterates. *Reading & Writing: An Interdisciplinary Journal*, *24*, 679–695. doi:10.1007/s11145-010-9256-9

Reas, C., & Fry, B. (2010). *Getting started with processing*. Sebastopol, CA: Maker Media.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536. doi:10.1038/323533a0

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. doi:10.1016/j.neunet.2014.09.003

Seidenberg, M. S. (2006). Connectionist models of reading. In G. Gareth (Ed.), *The Oxford handbook of psycholinguistics* (pp. 235–25). New York, NY: Oxford University Press.

Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, B. Miller., J. Ren., & O. Tzeng (Eds.), *Across languages: Orthography and the gene–brain–behavior link* (pp. 146–168). Baltimore, MD: Brookes.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568. doi:10.1037/0033-295X.96.4.523

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174. doi:10.1348/000712603321661859

Shen, H. H., & Ke, C. (2007). Radical awareness and word acquisition among nonnative learners of Chinese. *The Modern Language Journal*, *91*, 97–111. doi:10.1111/modl.2007.91.issue-1

Shu, H., Chen, X., Anderson, R., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development*, *74*, 27–47. doi:10.1111/cdev.2003.74.issue-1

Taft, M. (2006). Processing of characters by native Chinese readers. In P. Li., E. Bates., L. H. Tan., & O. J. L. Tzeng (Eds.), *Handbook of East Asian psycholinguisitcs* (Vol. 1, pp. 237–249). Cambridge, UK: Cambridge University Press.

Tamaoka, K., & Kiyama, S. (2013). The effects of visual complexity for Japanese kanji processing with high and low frequencies. *Reading and Writing: An Interdisciplinary Journal*, *26*, 205–223. doi:10.1007/s11145-012-9363-x

Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., & Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proceedings of the National Academy of Sciences, USA*, *102*, 8781–8785. doi:10.1073/pnas.0503523102

Treiman, R., & Kessler, B. (2011). Similarities among the shapes of writing and their effects on learning. *Written Language and Literacy*, *14*, 39–57. doi:10.1075/wll.14.1.03tre

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, *27*, 92–114.

Watson, A. (2012). Perimetric complexity of binary digital images. *The Mathematica Journal*, *14*, 1–41. doi:10.3888/tmj.14-5

White, T. G., Graves, M. F., & Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology*, *82*, 281–290. doi:10.1037/0022-0663.82.2.281

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, *8*, 221–243. doi:10.3758/BF03196158

Winskel, H. (2009). Reading in Thai: The case of misaligned vowels. *Reading and Writing: An interdisciplinary Journal*, *22*, 1–24. doi:10.1007/s11145-007-9100-z

Woollams, A., Lambon Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*, *114*, 316–339. doi:10.1037/0033-295X.114.2.316

Wu, N., Zhou, X., & Shu, H. (1999). Sublexical processing in reading Chinese: A development study. *Language & Cognitive Processes*, *14*, 503–524.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95. doi:10.1038/nature04262

Yang, J. F., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language*, *61*, 238–257. doi:10.1016/j.jml.2009.05.001

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, *54*, 145–160. doi:10.1016/j.jml.2005.08.002

Zorzi, M., Testolin, A., & Stoianov, I. P. (2014). Modeling language and cognition with deep unsupervised learning: A tutorial overview. *Frontiers in Psychology*, *4*(515), 2–14. doi:10.3389/fpsyg.2013.00515

## Appendix A

### The visual orthography measure

The visual orthography measure (VOM; Chang et al., 2015) was developed to quantify the visual complexity of any grapheme in any orthography. This measurement system includes four dimensions, each of whose strength in capturing different visual properties of graphemes has been demonstrated in prior reading research: (a) perimetric complexity: sensitive to ratio of written form to surrounding white space (Pelli et al., 2006), (b) number of simple features: sensitive to "building blocks" or strokes which graphemes comprise (Wu, Zhou, & Shu, 1999), (c) number of connected points: sensitive to continuity (Lanthier et al., 2009), and (d) number of disconnected components: sensitive to discontinuity (Gibson, 1969). Operational definitions of these four dimensions and their examples from five writing systems are provided next:

Perimetric complexity (PC): PC is defined as the square of the sum of the inside and outside perimeters of a grapheme divided by the product of $4\pi$ and the foreground area (Pelli et al., 2006; Watson, 2011). For example, in a 500-pixel × 500-pixel bitmap, 1 represents "ink" and 0 represents "paper"; if uppercase W has a 4,656-pixel perimeter and 136,602-pixel squared area, its perimetric complexity is 12.6287 (= 4656 × 4656/ 136602 /4π).

Number of simple features (SF): An SF is a discrete element of an image that can be discriminated independently from other features (Pelli et al., 2006). Namely, an SF is a mark drawn in a single movement in a specific orthography so that an SF can be a line, a dot, a curve, or a circle. For example, English < L >, Hebrew < ד >, Cree < ᑭ >, Telugu < ల >, and Chinese <人> each have two SFs.

Number of connected points (CP): A CP (or a junction) is an adjoining of at least two simple features. For example, English < F >, Hebrew < צ >, Cree < ᐳ >, Telugu < అ >, and Chinese <仆> each have two CPs.

Number of disconnected components (DC): Counter to the CP dimension, a DC is a simple feature or features in a set that do not adjoin any other feature. For example, English < i >, Hebrew < ה >, Cree < ᐄ >, Telugu < ఠ >, and Chinese <云> each have two DCs.

These four dimensions give objective, quantitative, and size-invariant estimations about grapheme complexity. Table A1 shows how these four dimensions capture different properties of five graphemes in five writing systems.

### VOM methods: To quantify grapheme complexity

From five writing systems, 131 orthographies were selected that have been specifically examined in cross-writing-system (Changizi & Shimojo, 2005), cross-alphabet (Seymour et al., 2003), and cross-Chinese-orthography (Chen, Chang, Chiou, Sung, & Chang, 2011) studies. *Ager's Omniglot: The Online Encyclopedia of Writing Systems and Languages* (Ager, 1998), the same source used in Changizi et al.'s (2005) study, was consulted to discern writing system classification for these orthographies, as well as the number of graphemes—21,821 graphemes in total. The programming

**Table A1.** Properties of graphemes in five writing systems.

| Writing system | Abjad | Alphabet | Syllabary | Alphasyllabary | Morphosyllabary |
|---|---|---|---|---|---|
| Orthography | Hebrew | Russian | Cree | Telugu | Chinese |
| Example grapheme | בּ | З | ᐄᐧ | ఞః | 面 |
| PC | 6.02 | 7.83 | 12.04 | 18.06 | 20.85 |
| No. of SF | 3 | 2 | 6 | 5 | 9 |
| No. of CP | 1 | 1 | 3 | 2 | 14 |
| No. of DC | 2 | 1 | 3 | 3 | 1 |

*Note*. PC = perimetric complexity; SF = simple features; CP = connected points; DC = disconnected components.

**Table A2.** Step-wise multiple regression of graphemic complexity measure components against graphemic inventory size

| | Model 1 | | | Model 2 | | | Model 3 | | | Model 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *SE B* | β | *B* | *SE B* | β | *B* | *SE B* | β | *B* | *SE B* | β |
| PC | 94.98 | 10.79 | .61** | 3.59 | 13.48 | .02 | −24.31 | 12.42 | −.16 | .65 | 10.35 | .01— |
| DC | | | | 1087.74 | 124.00 | .76** | 795.43 | 116.15 | .56** | 1185.76 | 103.48 | .83** |
| CP | | | | | | | 244.38 | 36.99 | .48** | 796.41 | 71.12 | 1.57** |
| SF | | | | | | | | | | −702.71 | 82.34 | −1.46** |
| $R^2$ | | | .38 | | | .61 | | | .71 | | | .82— |
| $R^2$ change | | | | | | .23 | | | .10 | | | .11— |
| *F* for change in $R^2$ | 77.49 | | | | 100.03** | | | 103.45** | | | 139.69** | |

*Note*. PC = perimetric complexity; DC = number of disconnected components; CP = number of connected points; SF = number of simple features.
$^{**}p < .01$.

language *Processing* (Reas & Fry, 2010) was used to construct a simple image of each grapheme in black Arial font against a 500 × 500-pixel white background. The visual complexity of a grapheme can be represented in two ways: (a) raw scores from each of the four dimensions of the visual orthography measure, and (b) a standardized composite score for representing overall complexity (calculated by transforming raw scores to within-dimension *z* scores and then averaging the resulting scores).

### *VOM implications: To predict grapheme inventory size predication*

To determine which weighted combination of dimensions could best predict grapheme inventory size across 131 orthographies, a multiple regression with four dimensions as predictors (i.e., mean scores from four dimensions for each orthography) was performed. The PC dimension was entered first given its reported significance in comparing grapheme complexity across orthographies (Pelli et al., 2006). Next, the three other dimensions were entered in a stepwise manner to determine whether any of them could account for remaining variance, above and beyond that explained by PC. The resulting model (i.e., Model 4 which included all four dimensions as significant predictors; $R^2$ = .82, $p < .01$) suggested that the four dimensions collectively, as opposed to the PC dimension alone, can best predict grapheme inventory size.

## Appendix B

### *A full list of orthographies*

**Table B1.** One hundred thirty-one orthographies (listed in alphabetic order) in five writing systems with overall complexity (Chang et al., 2015) and networks' learning epochs (mean of 10 runs).

| Writing system classification | Orthography | Grapheme inventory | Overall complexity (in *z* score) | Learning epochs (mean) |
|---|---|---|---|---|
| Alphabet | Albanian (Elbasan) | 40 | −0.54 | 15140 |
| | Albanian (Todhri) | 53 | −0.39 | 20820 |
| | Armenian (Eastern) | 38 | −0.55 | 32980 |
| | Asomtavruli | 38 | −0.43 | 21760 |
| | Avestan | 54 | −0.23 | 32390 |
| | Bassa | 30 | −0.55 | 21540 |
| | Belarusian | 32 | −0.43 | 37760 |
| | Bosnian | 30 | −0.40 | 34230 |
| | Bulgaria | 30 | −0.44 | 37890 |
| | Celtiberian | 28 | −0.28 | 15070 |
| | Cyrillic (Abkhaz) | 56 | 0.01 | 37640 |
| | Danish | 29 | −0.57 | 32570 |
| | Deseret | 38 | −0.58 | 28120 |
| | Dutch | 26 | −0.64 | 25900 |
| | English | 26 | −0.50 | 25900 |
| | Enochian | 22 | −0.49 | 15770 |
| | Finnish | 28 | −0.56 | 26470 |
| | Fraser | 40 | −0.55 | 16260 |
| | French | 26 | −0.65 | 25900 |
| | Glagolitic | 42 | 0.44 | 37580 |
| | Gothic (Wulfila) | 25 | −0.64 | 15660 |
| | Greek | 24 | −0.62 | 22020 |
| | German | 26 | −0.64 | 25900 |
| | Hungarian Runes | 46 | −0.12 | 27550 |
| | Icelandic | 32 | −0.49 | 32300 |
| | Italian | 21 | −0.66 | 25090 |
| | Kazakh | 42 | −0.43 | 43790 |
| | Korean (Hangeul) | 40 | 0.15 | 25240 |
| | Kyrgyz | 36 | −0.40 | 41190 |
| | Latin (ancient) | 21 | −0.56 | 16410 |
| | Latin (modern) | 41 | −0.28 | 25200 |
| | Macedonian | 31 | −0.45 | 29740 |
| | Marsiliana | 26 | −0.29 | 10880 |
| | Mkhedruli | 38 | −0.61 | 22060 |
| | Mongolian | 35 | −0.40 | 41310 |
| | Montenegrin | 33 | −0.39 | 34430 |
| | N'Ko | 27 | −0.52 | 15000 |
| | Norwegian | 29 | −0.57 | 32580 |
| | Nuskhuri | 38 | 0.16 | 17040 |
| | Old Church Slavonic | 45 | −0.22 | 33910 |
| | OldPermic (Abur) | 38 | −0.21 | 12540 |
| | Pahawh Hmong | 166 | 0.26 | 25850 |
| | Pollard Miao | 85 | −0.31 | 17500 |
| | Portuguese | 26 | −0.64 | 25910 |
| | Romanian | 31 | −0.56 | 26930 |
| | Runic (Danish Futhark) | 16 | −0.53 | 15000 |
| | Runic (Elder Futhark) | 24 | −0.31 | 18820 |
| | Russian | 33 | −0.32 | 40190 |
| | Santali (OlCemet') | 30 | −0.15 | 15510 |
| | Serbian | 30 | −0.45 | 30920 |
| | Somali (Osmanya) | 30 | −0.38 | 21520 |
| | Sorang Sompeng | 24 | 0.07 | 15000 |
| | Spanish | 27 | −0.61 | 27830 |
| | Swedish | 29 | −0.54 | 29350 |
| | Tajik | 35 | −0.40 | 39130 |
| | Theban | 25 | 0.23 | 19700 |
| | Ukrainian | 33 | −0.45 | 39690 |

(*Continued*)

**Table B1.** (Continued).

| Writing system classification | Orthography | Grapheme inventory | Overall complexity (in *z* score) | Learning epochs (mean) |
|---|---|---|---|---|
| | VarangKshiti | 30 | −0.40 | 15410 |
| | Yupik | 44 | −0.25 | 32120 |
| | Zhuyin Fuhao | 37 | −0.11 | 25250 |
| | Ancient Berber | 25 | 0.28 | 13320 |
| Abjad | Arabic | 28 | −0.26 | 29340 |
| | Aramaic (Early) | 22 | −0.45 | 15000 |
| | Hebrew | 32 | −0.58 | 18960 |
| | Middle Persian | 22 | −0.64 | 15000 |
| | Nabataean | 22 | −0.57 | 15000 |
| | Neo Tifinagh | 33 | −0.20 | 20330 |
| | Parthian | 22 | −0.60 | 15000 |
| | Pashto | 40 | 0.05 | 25100 |
| | Phoenician | 22 | −0.39 | 16110 |
| | Psalter | 21 | −0.75 | 11510 |
| | Sabaean | 29 | −0.28 | 15000 |
| | Samaritan | 22 | −0.08 | 20000 |
| | South Arabian | 28 | −0.31 | 15000 |
| | Syriac | 22 | −0.60 | 19230 |
| | Tifinagh | 33 | −0.16 | 17700 |
| | 'Phags-pa | 41 | 0.37 | 29310 |
| | Ahom | 45 | 0.02 | 27740 |
| | Amharic | 282 | −0.23 | 50900 |
| | Balinese | 84 | 0.85 | 37960 |
| | Batak (KaraBatak) | 32 | −0.70 | 15000 |
| | Bengali | 57 | 0.71 | 37860 |
| | Brahmi | 52 | −0.62 | 15690 |
| | Buhid (Mangyan) | 48 | 0.22 | 17890 |
| | Burmese | 62 | 0.23 | 29690 |
| | Dehong | 30 | −0.51 | 11310 |
| Alphasyllabary | Devanagari | 62 | 0.01 | 32810 |
| | Dives Akuru | 46 | −0.26 | 27370 |
| | Ethiopic (Ge'ez) | 234 | −0.29 | 54390 |
| | Gujarati | 64 | −0.34 | 34820 |
| | Gurmukhi | 60 | 0.32 | 31270 |
| | Hanuno'o (Mangyan) | 48 | 0.19 | 21790 |
| | Hindi | 66 | −0.02 | 36040 |
| | Inuktitut | 112 | −0.29 | 32960 |
| | Kannada | 50 | 0.17 | 28960 |
| | Kharosthi | 39 | −0.54 | 26420 |
| | Khmer | 130 | 1.12 | 43490 |
| | Lao | 78 | 0.51 | 27310 |
| | Lepcha Rong | 77 | −0.11 | 38060 |
| | Limbu | 45 | −0.29 | 22530 |
| | Malayalam | 69 | 0.03 | 33030 |
| | Manipuri | 57 | 0.26 | 32340 |
| | Marathi | 65 | 0.06 | 22330 |
| | Meroitic | 23 | −0.21 | 15000 |
| | Oriya | 66 | 0.12 | 32540 |
| | Redjang (Kaganga) | 36 | −0.46 | 17720 |
| | Sindhi | 51 | 0.26 | 25940 |
| | Sinhala | 71 | 0.45 | 28500 |
| | Soyombo | 86 | 1.10 | 23070 |
| | Syloti-Nagri | 38 | 0.23 | 16780 |
| | Tagalog | 45 | 0.18 | 29230 |
| | Tagbanwa | 42 | 0.21 | 18770 |
| | Tamil | 47 | 0.40 | 19360 |
| | Telugu | 70 | 0.07 | 27140 |
| | Thaana | 49 | −0.18 | 25240 |
| | Thai | 102 | 1.07 | 40250 |
| | Tibetan | 34 | 0.20 | 21700 |
| | Carrier | 195 | 0.10 | 25790 |
| | Cherokee | 85 | −0.49 | 25910 |

(*Continued*)

**Table B1.** (Continued).

| Writing system classification | Orthography | Grapheme inventory | Overall complexity (in *z* score) | Learning epochs (mean) |
|---|---|---|---|---|
| | Cree (Woodland) | 80 | −0.38 | 36550 |
| | Cypriot | 55 | 0.15 | 30900 |
| | Japanese (Hiragana) | 48 | 0.73 | 40030 |
| Syllabary | Japanese (Katakana) | 48 | 0.06 | 32550 |
| | Kpelle | 86 | 2.44 | 33710 |
| | LinearB | 71 | 1.66 | 49430 |
| | Ndjuka' | 57 | −0.31 | 25000 |
| | Ojibwe | 88 | −0.47 | 36020 |
| | Vai | 208 | 0.59 | 48120 |
| | Japanese (Kanji) | 2136 | 1.62 | 107350 |
| Morphosyllabary | Chinese (Simpified) | 2707 | 3.22 | 63770 |
| | Chinese (Traditional) | 2707 | 3.79 | 194790 |

## Appendix C

### *The behavioral experiment (Chang, 2014)*

#### *Design*

A 8 × 6 mixed design was used with participants' first-language (L1) orthography as a between-participants factor (Hebrew, English, Russian, Arabic, Hindi, Telugu, Japanese, and Chinese participants) and grapheme group as a within-participants factor (Hebrew, Russian, Cree, Telugu, simple Chinese characters, and complex Chinese characters).

#### *Participants*

A total of 480 participants (60 participants for each of eight L1 orthographies) took part in the experiment. Across all eight orthographies, the participants were matched by age ($M$ = 26.88, $SD$ = 5.16), $F < 1$, and met the following criteria: (a) native speaker of one of the languages in eight target orthographies, (b) age from 18 to 35, and (c) no vision or hearing impairments.

#### *Same-different judgment task*

This task tapped individuals' perceptual processing of graphemes. Each trial began with a black fixation cross appearing for 300 ms, followed by a pair of graphemes appearing for up to 1,000 ms, followed by a blank for 1,000 ms. The participants were instructed to judge whether two graphemes were the same or different using their index fingers; response keys were counterbalanced across the four stimulus lists. After instructions, the participants were given 12 example trials with answers, 36 practice trials without feedback, and 360 critical trials with randomized presentation.

#### *Stimuli*

The stimuli comprised six grapheme groups of increasing complexity (i.e., Hebrew, Russian, Cree, Telugu, simple Chinese, and complex Chinese). For the same-different judgments, graphemes paired with themselves comprised "same" pairs; all graphemes in each orthography (except for Chinese) were exhaustively used. For "different" pairs, graphemes were matched by complexity, with individual graphemes appearing only once during testing. For practical reasons, not all possible pairs of graphemes were used. Four lists were created to allow generalization of results to other grapheme combinations. Within each list, complexity varied by grapheme group according to the following ranking (overall complexity of grapheme pairs per orthography across all four lists): Hebrew (−0.58) < Russian (−0.32) < Cree (−0.32) < Telugu (0.07) < simple Chinese (0.48) < complex Chinese (3.79), $F(5, 1439) = 2339.61$, $p < .001$. Between lists, no complexity differences in grapheme pairs were found for any grapheme group, $F(3, 1439) = 1.64$, $p = .18$. Each list contained 360 pairs.

**Table C1.** Means and standard deviations (in parentheses) of proportion accuracy in examining whether L1 orthographic complexity differentially affects perceptual variability across grapheme complexity levels.

| Participant (ordered by L1 orthographies from simple to complex) | Grapheme group (ranked complexity from simple to complex): 1 Hebrew | 2 Russian | 3 Cree | 4 Telugu | 5 Simple Chinese | 6 Complex Chinese | *F* | $\eta_p^2$ | Pairwise comparison (With bonferroni adjustments) |
|---|---|---|---|---|---|---|---|---|---|
| Hebrew (–.58) | 0.93 (0.05) | 0.93 (0.06) | 0.91 (0.05) | 0.79 (0.10) | 0.79 (0.11) | 0.48 (0.15) | 329.17 ** | .848 | 1,2,3 > 4,5 > 6 |
| English (–.50) | 0.87 (0.08) | 0.88 (0.08) | 0.87 (0.08) | 0.74 (0.13) | 0.77 (0.11) | 0.45 (0.16) | 321.22 ** | .845 | 1,2,3 > 4,5 > 6 |
| Russian (–.32) | 0.89 (0.11) | 0.90 (0.08) | 0.89 (0.09) | 0.78 (0.13) | 0.78 (0.12) | 0.49 (0.16) | 255.89 ** | .813 | 1,2,3 > 4,5 > 6 |
| Arabic (–.26) | 0.85 (0.12) | 0.88 (0.13) | 0.86 (0.13) | 0.75 (0.12) | 0.75 (0.14) | 0.48 (0.14) | 311.27 ** | .841 | 1,2,3 > 4,5 > 6 |
| Hindi (–.02) | 0.86 (0.09) | 0.87 (0.09) | 0.86 (0.08) | 0.74 (0.14) | 0.71 (0.12) | 0.45 (0.14) | 308.67 ** | .840 | 1,2,3 > 4,5 > 6 |
| Telugu (.07) | 0.86 (0.10) | 0.87 (0.11) | 0.84 (0.12) | 0.81 (0.12) | 0.73 (0.11) | 0.45 (0.11) | 377.34 ** | .865 | 1,2 > 3,4 > 5 > 6 |
| Japanese (1.62) | 0.91 (0.07) | 0.92 (0.06) | 0.92 (0.06) | 0.83 (0.10) | 0.87 (0.10) | 0.62 (0.13) | 209.87 ** | .781 | 1,2,3 > 4,5 > 6 |
| Chinese (3.79) | 0.90 (0.07) | 0.91 (0.06) | 0.90 (0.08) | 0.81 (0.08) | 0.88 (0.08) | 0.69 (0.13) | 159.56 ** | .730 | 1,2,3 > 5 > 4 > 6 |

*Note*. For the pairwise comparison with Bonferroni adjustments, all *ps* < .001.
$^{**}p < .001$.

### *Results*

An 8 × 6 (L1 Background × Grapheme Complexity) analysis of variance was conducted with proportion accurate as the dependent measure. The main effect of L1 background was significant, $F(7, 472) = 12.31$, $p < .001$, $\eta_p^2 = .154$; the main effect of grapheme complexity was also significant, $F(5, 2360) = 2224.13$, $p < .001$, $\eta_p^2 = .825$. There was a significant interaction between L1 background and grapheme complexity, $F(35, 2360) = 16.26$, $p < .001$, $\eta_p^2 = .194$. Pairwise comparisons using Bonferroni adjustments are provided in Table C1 along with the means and standard deviations of proportion accurate.

## Appendix D

**Table D1.** An illustration of graphemes with varying complexity for eight orthographies.

| Orthographies ranked by overall complexity (average grapheme complexity; in *z* score) | Graphemes with varying complexity within Its orthography (sampled by extreme values and quartile): Min. | $Q_1$ | $Q_2$ | $Q_3$ | Max. |
|---|---|---|---|---|---|
| 1 Hebrew | י | ג | ף | א | ב |
| 2 English | I | L | D | A | W |
| 3 Russian | Г | Р | И | В | Ё |
| 4 Arabic | َ | د | ذ | ت | ش |
| 5 Hindi | ं | त | थ | छ | अँ |
| 6 Telugu | ు | ూ | ణ | త | ఫ |
| 7 Japanese | 一 | 拉 | 炭 | 暫 | 鬱 |
| 8 Chinese | 一 | 欣 | 挽 | 惃 | 纛 |

*Note*. min. = minimum; $Q_1$ = first quartile; $Q_2$ = second quartile; $Q_3$ = third quantile; max. = maximum.

## Appendix E

### Additional modeling details

Simulations were carried out using a modified version of the Lens simulation software package developed by Doug Rohde (http://tedlab.mit.edu/~dr/Lens/). The modified software and all necessary network scripts and example files are available for download at http://www.cnbc.cmu.edu/~plaut/ChangPlautPerfetti15SSR.

### Network architecture

Details of the network architecture are included in the main text. Bias connections for hidden and output units (from a unit with fixed activation of 1.0) were initialized to random values sampled uniformly within [−2.05, −1.95]. All other (nonbias) weights were initialized randomly, sampling uniformly within [−0.1, 0.1]. Ten different sets of initial random weights and biases were generated, corresponding to 10 instances or "participants." Each such instance was trained on each of the 131 orthographies, starting from its specific initial weights. Thus, exactly the same 10 networks and the same training procedures were applied to each of the 131 orthographies.

### Stimuli

Each 38 × 38 pixel grapheme image was converted to white-on-black and normalized so that pixel values ranged from 0.0 to 1.0 (inclusive). These values formed both the input activations and target output activations for the network.

### Training procedure

The network was trained with back-propagation using momentum descent (learning rate of 0.01, momentum of 0.8) to minimize cross-entropy error. Activations within 0.05 of their specified targets were considered correct and generated no error. Weights were updated after the presentation of each grapheme (i.e., online learning), each of which was presented once per training epoch. Training was halted once the average cross-entropy error per grapheme for a given epoch fell below 10.0.

### Testing procedure

To compare two graphemes in an approximation a same-different judgment task, the 635 hidden activations generated by each grapheme were converted back to the net inputs that generated them by applying the inverse of the sigmoidal input–output unit function. These two sets of net inputs were then corrupted with Gaussian noise ($M$ = 0.0, $SD$ = 1.0), converted back to activations by applying the sigmoid function, and then compared by computing 1.0 – $r$, where $r$ is their Pearson correlation. This gives a measure of the dissimilarity of the two representations. This measure was computed 100 times (for different noise samples) and averaged. Greater representational dissimilarity was assumed to correspond to better performance on "different" trials, and poorer performance on "same" trials.