Running Head: HETEROGENEITY IN READING

INDIVIDUAL DIFFERENCES IN THE NEURAL AND COGNITIVE MECHANISMS OF SINGLE WORD READING

Simon Fischer-Baum[1*], Jeong Hwan Kook[2], Yoseph Lee[1], Aurora Ramos-Nuñez[1,3] & Marina Vannucci[2]

[1] Department of Psychology, Rice University, P.O. Box 1892. Houston, TX 77251

[2] Department of Statistics, Rice University, P.O. Box 1892. Houston, TX 77251

[3] Department of Social Sciences, Coastal College of Georgia, One College Drive, Brunswick, GA 31520

*Corresponding author

Department of Psychology, MS-25

Rice University

P.O. Box 1892

Houston, TX 77251

Tel: 713-348-3841

simon.j.fischer-baum@rice.edu

Abstract

Written language is a human invention that our brains did not evolve for. Yet, most research has focused on finding a single theory of reading, identifying the common set of cognitive and neural processes shared across individuals, neglecting individual differences. In contrast, we investigated variation in single word reading. Using a novel statistical method for analyzing heterogeneity in multi-subject task-based fMRI, we clustered readers based on their brain's response to written stimuli. Separate behavioral testing and neuroimaging analysis shows that these clusters differed in the role of the sublexical pathway in processing written language, but not in reading skill. Taken together, these results suggest that individuals vary in the cognitive and neural mechanisms involved in word reading. In general, neurocognitive theories need to account not only for what tends to be true of the population, but also the types of variation that exist, even within a neurotypical population.

Key Words: fMRI, visual word recognition, reading aloud, sublexical processing, individual differences

Cognitive scientists typically assume individuals rely on a common set of cognitive operations with essentially the same neural organization (De Schotten and Shallice, 2017). At the same time, there is clear variability across individuals, in terms of skills, experiences and abilities. Understanding both general patterns of cognitive and brain functioning and the ways that individual subjects depart from those general patterns is critical for understanding how the human mind works.

The current project focuses on variability in the cognitive and neural processes involved in reading. Reading is a particularly fruitful topic for investigating neuro-variability. Written language is a relatively recent human invention, and therefore reading is not something that our brains evolved to do. Unlike spoken language, literacy skills are learned through explicit instruction, with different learning techniques that have been proposed (Rayner et al., 2002). Yet, individuals from a wide variety of linguistic and educational backgrounds have been argued to rely on essentially the same reading system, both in terms of cognitive architectures (Perfetti, 2011) and neural substrates (Rueckl et al., 2015). This extends even to readers of vastly different writing systems, with similar brain regions involved in processing logographic languages, transparent and opaque alphabetic languages and even Braille which is read by touch rather than by sight (Rueckl et al., 2015; Reich et al., 2011). Despite the fact that reading is a taught skill that our brains did not evolve for, it appears that most people learn to do it approximately the same way.

At the same time, it is clear that readers vary, at least to some degree, in how they process written language. Even among highly skilled, literate adult readers without any evidence of reading difficulties like dyslexia, there are differences in reading skill

(Andrews, 2012). These differences in reading skills have been mapped onto variability in the efficacy of different cognitive processes, like linking from letters to sound (Gough and Tunmer, 1986), recognizing familiar words (Andrews, 2012; Perfetti, 2007), comprehending spoken sentences (Gough & Tunmer, 1986) and knowledge about the phonological structure of words (Johnson & Kirby, 2006). Variability in reading skill has been linked to neural differences, for example with differences in sight word and phonemic decoding efficiency relating to different degrees of BOLD response to written words in temporal, occipital and parietal regions during functional MRI (Welcome and Joanisse, 2012), or differences in the semantic influences on reading relating to differences in the integrity of white matter pathways connecting regions within the left temporal lobe and between the superior temporal gyrus and angular gyrus (Graves et al., 2014 see also Jobard et al., 2011, Seghier et al., 2008, 2011, Kherif et al., 2009, Welcome and Joanisse, 2014).

How does this variability relate to our understanding of how we read words? Most theories of word reading assume that there are multiple pathways for reading words aloud, with the lexical/semantic pathway focusing on processing the meaning and pronunciation associated with familiar words and a sublexical/phonological pathway with a sounding out procedure that maps directly from letters to sounds irrespective of whether or not the stimulus is a known word (Coltheart et al., 2001; Plaut et al., 1996). While both pathways are capable of generating correct word pronunciations, at least for familiar words with regular spelling to sound correspondences, these theories also assume that when readers are confronted with a word, both pathways are activated simultaneously and in parallel. However, just because both pathways are used to read

does not mean that they carry equal weight. A variety of studies have demonstrated that the relative weight of the lexical and sublexical pathways can be manipulated strategically based on the makeup of words in the experiment (Zevin and Balota, 2000; Monsell et al., 1992; Rastle and Coltheart, 1999 cf. Lupker, Brown and Colombo, 1997) or based on task demands (e.g. Bitan et al., 2005). Some researchers have proposed that the weight of these routes also varies as a function of individuals, with some participants depending more on lexical processes when they read and others depending more on sublexical processes (Baron and Strawson, 1976; Woollams et al., 2016), though this position remains a matter of debate (Yap et al., 2012; Brown, Lupker and Colombo, 1994). We explore that possibility here.

Our approach involves a combination of functional magnetic resonance imaging (fMRI), behavioral experiments designed to tap into phonological processing of written words, and standardized tests of reading skill. We take a novel, data-driven approach to analyzing multi-subject, task-based fMRI data (Zhang et al., 2014, 2016) that can cluster subjects into subgroups characterized by similar patterns of brain responses across the whole brain to written words. We then ask whether these subgroups differ in the engagement of sublexical and/or lexical processes on the basis of their performance on the behavioral tasks and on analyses of an orthogonal set of neuroimaging data. Previous studies have demonstrated that clustering techniques can be used to identify subgroups of readers on the basis of the neuronal activation for reading aloud (Kherif et al., 2009). However, these previous experiments collected only limited behavioral data, making it is difficult to interpret how these subgroups differ cognitively.

The current study involves a much richer set of behavioral and neuroimaging measures designed to evaluate lexical and sublexical processing. Behavioral tasks include classic effects in both lexical decision and reading aloud. The *pseudohomophone lexical decision* paradigm can be used to test phonological processing in the context of a lexical decision task. In this task, pseudowords are evenly divided between pseudohomophones, or pseudowords that are prounounced like real words (e.g. BRANE) and non-pseudohomophones, or matched pseudowords whose pronunciations do not correspond to familiar words (e.g. BRAME; Besner and Davelaar, 1983; Rubenstein, Lewis and Rubenstein, 1971). Participants are slower and make more errors when they have to reject pseudohomophones than non-pseudohomophones, which has been used to argue for an influence of sublexical processing on a task that could be carried out exclusively with lexical/semantic processing. In reading aloud experiments, researchers analyze the speed with which participants read different types of orthographic stimuli. The *lexicality* effect, or the difference in the speed with which pseudowords and familiar words are read, has been used to isolate sublexical processing ability, with smaller lexicality effects indicating stronger sublexical processing. The *regularity* effect, or the difference in the speed with which regular words (e.g. CLAM), that can be correctly sounded out by the sublexical route, and irregular words (e.g. YACHT), with idiosyncratic pronunciations that can only be correctly processed by the lexical route, are read has been used to isolate lexical processing ability, with smaller differences between these two word types indicating a greater reliance lexical/semantic processing. The size of the pseudohomophone effect, lexicality effect and regularity effect all vary, to some extent, by participant strategies

(e.g. Wagenmakers et al., 2008; Coltheart and Rastle, 1994), but individual difference approaches have found clear relationships in performance on the two tasks, as well as clear links to individual variation in reading skill (Katz et al., 2012).

Neuroimaging has also been used to evaluate lexical and sublexical processing. A neurobiological distinction has been made between dorsal and ventral reading routes. Dorsal reading areas, including the parietal lobe and superior temporal gyrus have been linked to sublexical/phonological processing of written words, while ventral reading areas including ventral occipitotemporal and middle temporal gyrus regions have been linked to lexical/semantic processing (Schlaggar & McCandliss, 2007). In a recent meta-analysis, Taylor and colleagues (2013) argue that lexical and sublexical processes can be partially identified by comparing the BOLD response to word and pseudoword stimuli, with greater responses to pseudowords indicating sublexical engagement and greater response to words indicating lexical engagement. Using this approach, they identified dorsal regions, specifically in the inferior parietal lobule and inferior frontal gyrus, as critical nodes in the sublexical reading pathway, and ventral regions, specifically the anterior fusiform and middle temporal gyrus, as critical nodes in the lexical reading pathway. The angular gyrus was also highlighted as part of the lexical pathway.

An alternative approach to evaluating the activation of semantic and phonological representations in different brain regions during reading tasks is the use of representational similarity analysis (RSA) fMRI (Fischer-Baum et al., 2017; Zhao et al., 2016). This approach can identify the type of information about written words represented in different cortical regions on the basis of the similarity of the fine-grained

patterns of activity to different words. Consider the word DOUGH. DOUGH is related to TOUGH visually, SEW phonologically, and BREAD semantically. According to RSA logic, brain regions that have similar patterns of activity for DOUGH and SEW process phonology, rather than semantics or visual information while regions that have similar patterns of activity for DOUGH and BREAD process semantics.

Our study focuses on a sample of college students enrolled at a highly selective, four-year university who have no previous or current diagnosis of dyslexia, dysgraphia or other learning difference. This sample is not representative of variation in reading ability across the entire population, as our participants are likely extremely skilled readers. To assess this, participants underwent a series of standardized measures of reading skill; the Nelson-Denny Reading Test (Form G; Brown, Fischo and Hanna, 1993), which was used to examine vocabulary knowledge, reading comprehension skill and reading rate and the Test of Word Reading Efficiency (TOWRE-2; Torgensen, Wagner and Rashotte, 2012), which was used to examine decoding and sight-word reading ability. We focus on this population because neurotypical, English-speaking, college students are the most commonly studied population in the development of models of reading, and are therefore the population that is typically assumed to be homogenous with respect to the processes involved in reading. It is of interest, therefore, to investigate whether there is variation in reading processes even within this population.

The study was divided into two sessions. In one session, participants read a combination of words and pseudowords aloud while undergoing fMRI scanning. In another session, participants were given a battery of behavioral tasks; standardized

measures of reading skill (reading comprehension, reading rate, vocabulary knowledge, decoding ability), a pseudohomophone lexical decision study and a reading aloud experiment with pseudowords, regular words and irregular words. In order to avoid circular fMRI analyses (Kriegeskorte et al., 2009; Vul and Pashler, 2012), only data from the first run of the fMRI session was used to cluster participants into subgroups of readers on the basis of how their brains respond to written words. Data from the remaining five runs of the fMRI session were then used to compare how the brains of these two clusters of participants respond to different types of orthographic stimuli, both a whole brain univariate analysis compared groups in their response to words versus pseudowords and a multivariate analysis using RSA to compare groups on how key regions of the reading network process semantic and phonological information about familiar words. Finally, these subgroups were compared in their performance on behavioral tasks, specifically the lexical decision task and the reading aloud task to determine if they differed with respect to size of their lexicality, regularity and pseudohomophone effect. Together, the neuroimaging and behavioral data suggest that the clustering algorithm identified two groups of reader that differ in how effortful the use of the sublexical/phonological route was in reading, despite no apparent differences in reading skill.

Methods

Participants. Thirty Rice University students (23 females; *M* = 19.73 years old; range 18-28 years) participated in the study. All were 18 years of age or older, Native English speaker, with no history of dyslexia, dysgraphia or neurological disorders and no contraindications to MRI. The participants were compensated $25 for each of the

neuroimaging and the behavioral testing sessions. This study was carried out in accordance with the recommendations of the Rice University Institutional Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Behavioral testing. Standardized measures of vocabulary knowledge, reading comprehension, reading rate and decoding ability were collected using the Nelson-Denny Reading Test and the TOWRE-2. The Nelson-Denny Reading Test is a standardized reading comprehension test for college level students and includes 80 multiple choice vocabulary questions and 38 reading comprehension questions. We also obtained a measure of reading rate (number of words read in the first minute of the reading comprehension section). For decoding ability, participants were administered two subtests of the TOWRE-2; Sight Word Efficiency (SWE), which measures the number of words that an individual can accurately identify within 45 seconds (108 maximum) and Phonemic Decoding Efficiency, which measures the number of pronounceable nonwords that an individual can accurately decode within 45 seconds (66 maximum). Participant's performance on both the Nelson-Denny and the TOWRE-2 were compared to normative data.

Each participant was also administered a pseudohomophone lexical decision experiment. Materials and methods were taken directly from Besner & Davelaar (1983). The task had a total of 117 items that were categorized into 3 groups presented in random order: words (n=39), nonwords that were pseudohomophones (n=39), and nonwords that are non-pseudohomophones (n=39), with a short practice session prior

to the experimental trials. Nonword stimuli were matched on length in both number of letters and number of syllables and for number of orthographic neighbors. We used a smaller set of word stimuli than is typical in these experiments to save some time in the extensive behavioral testing session, meaning that there were twice as many nonwords than words. This modification may have reduced the size of the pseudohomophone effect, but as we show below, the effect is still present with this design. The experiment was presented on DmDx (Forster and Forster, 2003). A trial consisted of a fixation cross for 1500ms followed by the experimental stimuli presented in uppercase, size 36 Times New Roman font. The word disappeared following 1500ms or once a response was produced. The participants pressed the Left Shift key to indicate that the stimulus was not a word and the Right Shift key to indicate that the stimulus was a word, and were instructed to respond as quickly and accurately as they could.

Finally, each participant was also administered a regular word, exception word and nonword reading experiment, also presented in DmDx. Materials were drawn from Baron and Strawson (1976), and included regular words (n=49), exception words (n=47) and pseudowords words (n=30) presented in a random order, with a short practice session prior to the experimental trials. These words are matched for length and number of orthographic neighbors, but not frequency, with exception words being significantly higher in frequency than the regular words. A trial consisted of a fixation cross for 1500ms followed by the experimental stimuli presented in uppercase, size 36 Times New Roman font. The word disappeared following 1500ms or once a response was produced. Responses were produced into a Cyber Acoustics USB Stereo Headset and Boom Mic, and reaction times were measured from the onset of the stimulus to the

triggering of the voice key. One participant had to be removed from the lexical decision study and one from the reading aloud study because of equipment malfunctioning.

For half of the participants, the behavioral session preceded the fMRI session, and for the other half the order was reversed. On average, the time between the behavioral and MRI session was 13.5 days, with a range from 0 (both sessions on the same day) to 48 days apart.

fMRI acquisition. A high-resolution T1-weighted structural and six task-based functional scans were acquired during an 80-minute session using a 3T Siemens Magnetom Tim Trio scanner equipped with a 32-channel head coil. Scanning was done at the Core for Advanced Magnetic Resonance Imaging (CAMRI), at Baylor College of Medicine. A T1-weighted structural scan was collected first, followed by six consecutive 9.5-minute functional scans. In between runs, subjects remained lying down in the scanner waiting for the next run to begin and were informed that the next run would begin shortly and to remain still. All 30 subjects participated in the imaging session. The T1-weighted structural scan involved the following parameters: TR=2600ms, TE=3.03ms, FoV=256mm, matrix size=256x256, voxel size = 1x1x1 mm^3. Functional runs were six 9.5-minute scans obtained by using the following sequences: TR = 2000ms, TE = 30ms, FoV = 200 mm, voxel size = 2x2x2 mm^3, slice thickness = 2mm. A total of 280 volumes per run each with 62 slices were acquired in the axial plane to cover the whole brain. The task-based runs involved an ungrouped event-related design. Thirty (30) orthographic stimuli (20 familiar words, and 10 pronounceable pseudowords, Appendix A) were presented once during each run, though participants were shown the same set of words across the 6 runs, with a different random order in

each run. The repetition across runs was necessary for the representational similarity analysis, as the approach compares the pattern of activation to individual words. However, we chose to cluster participants on the basis of the first run to avoid the effects of stimulus repetition on BOLD response. Stimuli were presented for 2 seconds, with a trial onset asynchrony of 12 to 20 seconds; this long ISI was selected on the basis of prior work with the clustering algorithm described below (Zhang et al., 2014, 2016), and pilot data that suggested that the clustering algorithm was not well suited for rapid event related designs. Participants read the words aloud while they were in the scanner, and responses were recorded for accuracy, though data was not recorded in such a way that reaction times could be calculated.

fMRI data analysis. Image preprocessing was conducted using SPM12 with the standard processing pipeline. The data were slice time corrected to correct for the difference in time between the first and the last slice acquired, realigned by removing motion artifacts, coregistered by overlaying the structural and functional images. For the clustering analysis and the univariate groupwise difference analysis, data was then normalized by warping to fit the standard the Montreal Neurological Institute template, and spatially smoothed using a Gaussian kernel of FWHM 4mm. For the representational similarity analysis, no normalization or smoothing was carried out.

Clustering Analysis. Time series data from the first experimental runs was used for the clustering analysis. A 3D parcellation of the data were performed using the MarsBaR toolbox in SPM 12. The Automatic Anatomical Labeling (AAL) brain atlas was used to obtain the parcellation, resulting in 90 ROIs, excluding the regions associated with the cerebellum. A Bayesian Nonparametric Spatio-Temporal Model for Multi-Subject fMRI

Data was applied. This method has been reported previously in the literature (Zhang et al., 2014, 2016). Methods are implemented in the Matlab GUI *NPBayes-fMRI* (Kook et al. 2017) processed data and software available at the GitHub address (https://github.com/rimehi/NPBayes_fmri, to be released publicly upon acceptance) that is comprised of two main interfaces, one for model fitting and one for the visualization of the results. Briefly, the clustering algorithm uses a unified, single stage Bayesian framework for the analysis of task-related brain activity in multi-subject fMRI data that eschews the traditional two-stage analysis which divides within subject analyses from between subject inferential statistics. Instead, the model specifically accounts for between-subject heterogeneity in BOLD response via a spatially informed multi-subject nonparametric variable selection prior. The model simultaneously estimates subject- and group-level statistical parametric maps of responses to different stimuli, with the subject level analysis borrowing strength in the estimation of the parameters from other subjects showing similar activation patterns. In this way, the model can cluster subjects into groups of individuals who show similar patterns of activation in response to specific kinds of stimuli. For model fitting, we used the default set of parameters provided in *NPBayes-fMRI*. Zhang et al. (2016) and Kook et al. (2017) describe the role of these parameters and offer general guidelines regarding the sensitivity of the results to the choice of their values. Zhang et al. (2016), in particular, show robustness of the results to the choice of several model parameters and only notice small sensitivity to the parameters that capture information on neighboring structure among ROIs. Here, we obtained a neighboring network by calculating Euclidean distances between pairs of ROIs using the coordinates defined in the Montreal Neurological Institute (MNI) space

and then thresholding the distances. We chose a threshold so that ROIs would have five neighbors on average. This information was used in the model to capture spatial correlation among neighboring ROIs when specifying an a priori probability of activation. Clustering of the subjects was consistent under small deviations of the prior specification from the default setting, with clustering configurations showing between 70% and 90% overlap. Activation maps were obtained by using a pre-specified Bayesian False Discovery Rate (BFDR) of 0.01, to control for multiple comparisons.

Between-Group Univariate Analysis. To ensure that the data used for clustering were independent of the validation data, preprocessed and normalized time series data from runs 2-6 were analyzed using a univariate approach. For the first-level analysis, data were entered into a subject specific, fixed-effect analysis using the general linear model, with one regressor based on the onset times for words and a second regressor based on the onset times for pseudowords (both modeled as events) deconvolved with a hemodynamic response function, along with six motion parameters and a regressor for scanner drift. For each subject, a t-map contrast of pseudowords vs. words was computed. These t-maps were then entered into a second level analysis that looked at whether the within subject pseudoword vs. word contrast differed as a function of subgroup assignment based on data from all stimuli from Run 1. Statistical thresholding was based on cluster-extent based thresholding, with a Gaussian Random Field cluster-size correction of FWE $p < .05$. An initial conservative threshold analysis with the primary $p < .001$ and the cluster-size correction $k = 42$ failed to identify any clusters (Woo, Krishnan and Wager, 2014), so a more liberal analysis with the primary $p < .005$ and the cluster-size correction $k = 95$ was applied. Both of these correspond to a FWE

k-extent threshold of $p < .05$. We then looked at the peak voxel of each cluster to determine the relationship between word and pseudoword activation for each subgroup to determine the source of the interaction.

Representational Similarity Analysis. For the RSA, no smoothing or normalization was applied during preprocessing. A general linear model predicting BOLD response that included the timing of each individual word (modeled as an event) deconvolved with a hemodynamic response function, six motion parameters and scanner drift was applied, resulting in beta-weights for each word in each run against fixation. We obtained 20 beta-weight maps for each subject, with each map reflecting the brain's response to each word in the experiment. Following methods from previous studies (Fischer-Baum et al., 2017; Rothlein and Rapp, 2014), these beta-weight maps were then mean centered within each subject. Anatomical ROI analyses, based on warping the AAL map to each participant's native space using the backward deformation fields generated from the SPM segmentation step, were applied to these 20 individual-word beta -weight maps for each participant. Specifically, we focused on four regions of interests: the left inferior parietal lobule, the left angular gyrus, a broad ROI in the left ventral occipitotemporal lobe that combined the fusiform gyrus, inferior temporal gyrus and inferior occipital gyrus, and an inferior frontal gyrus ROI that combined the pars triangularis and pars opercularis.

For each ROI and each participant, a vector of beta-weights for the voxels within that ROI was extracted for each of the 20 words, and a similarity matrix of word-to-word similarity for this region was calculated based on a tie-corrected Spearman correlation of the beta -weight vectors for each word to every other word. For each individual, the

similarity between brain-based and theoretical similarity matrices was calculated. Phonological similarity was estimated using the phonological edit distance function in Phonological CorpusTools (Hall et al., 2016). In this measure, the similarity of two forms is calculated by calculating the number of one-feature changes needed to transform one phonological sting into another. Semantic similarity was estimated using an online pairwise Latent Semantic Analysis (LSA) calculator (lsa.colorado.edu/), a technique from distributional semantics that analyzes meaning based on which words appear in similar contexts (Landauer, Foltz and Laham, 1998). The result of the correlation between the brain-based and predicted similarity structure is taken to indicate the degree to which the region is processing phonological and/or semantic information in this group of participants.

Two analyses were carried out with this approach. First, we looked at the whole sample, asking whether the average correlation was significantly greater than zero. Second, we look at whether there are differences in these measures on the basis of which subgroup cluster they get assigned to, with a separate measure of semantic and phonological information processing for all ROIs calculated for each group.

Results.

Overall Behavioral Results. The 30 participants in the study were all students at a highly competitive private university without a previous diagnosis of dyslexia. Table 1 reports descriptive statistics for the two standardized tests of reading skill. Unsurprisingly, their scores on standardized measures of reading skill are quite high. Based on the Nelson-Denny task, we found that they were in the 90th percentile for vocabulary knowledge, the 89th percentile for reading comprehension and the 72nd

percentile for reading rate. Based on the Test of Word Reading Efficiency (TOWRE-2), they were in the 85th percentile for sight word efficiency and the 85th percentile for phonemic decoding efficiency.

Table 2 reports the group average results of the pseudohomophone lexical decision and the reading aloud study. We replicated the standard finding in the literature, with participants taking 13ms longer to reject the pseudohomophones than matched pseudowords ($t(28) = 3.96$, $p = .0004$[1]). We also replicated classic reading aloud effects of both lexicality and regularity. Participants took 87ms longer to name pseudowords than words ($t(28) = 9.32$, $p < .0001$) and were 23ms slower with irregular words than with regular words ($t(28) = 3.88$, $p = .0005$). Correlations across participants between these behavioral measures are reported in Table 3.

The critical test in this study is whether there were between group differences on these effects (pseudohomophone, lexicality and regularity), when groups are defined based on clustering of the fMRI data. In the next section, we describe the results of the fMRI clustering algorithm, and subsequent fMRI analyses based on these clusters. Finally, we reanalyze the behavioral data on the basis of these subgroups.

fMRI Results.

Behavioral Results. Overall, participants were highly accurate in reading words and pseudowords aloud in the scanner (mean = 98.2% correct, SD = 3.1%). All participants participated in 6 runs of the task, with the same set of 30 words repeated in each run. Three of the participants made a large number of no response errors (>10) in a single run. The first reported that she forgot the instructions, the second that she dozed off for part of one run and the third that his glasses needed to be adjusted. These runs were

[1] Statistical tests for reaction time data were carried out after a log transformation.

excluded from subsequent analysis. Participants made the occasional error in decoding nonwords (<.5% of all nonword trials), and there was only a single instance of a word being pronounced incorrectly.

Clustering Analysis. The results of the model estimation are shown in **Figure 1**, based on a comparison of all words to a blank screen baseline, with Bayesian False Discovery Rate (BFDR) = .01, to control for multiple comparisons. The top of Figure 1 shows the cluster dendrogram indicating which subjects show a similar pattern of response to written words based on Automated Anatomical Labeling (AAL) regions. Based on the dendrogram obtained via hierarchical clustering on the thresholded beta coefficients, we select two as number of clusters. Measures commonly used for interpretation and validation of consistency within clusters of data, such as the within cluster sums of squares and the average silhouette, returned two as the optimal number of clusters with 10 subjects in Subgroup 1 and 20 subjects in Subgroup 2.

The bottom of Figure 1 shows the posterior group-level β-maps for these two subgroups at six axial slices, as well as the difference in the βs between the two subgroups. The Bayesian framework of the clustering algorithm means that it would be inappropriate to do traditional frequentist statistics to determine which regions differ significantly across groups. However, we can determine which AAL regions show large β-weights (>1) in each group, and which AAL regions have a numerical difference in *a posteriori* β-weights response across groups. These regions are compared to the Taylor et al. (2013) meta-analysis neuroimaging findings of the neural substrates of reading aloud.

Both subgroups showed increased activation to words over baseline in bilateral occipitotemporal regions involved in visual processing of written words (inferior Occipital Sulcus, Lingual Gyrus, as well as the Left Fusiform Gyrus), and bilaterally in the superior temporal lobe (Superior Temporal Gyrus, Heschl's Gyrus), which is involved in phonological processing. Most notably, the two groups appear to differ by degree rather than by double dissociation. The second subgroup showed higher β values than the first subgroup across the cortex. Regions that had high β-values for the second subgroup and low β-values for the first subgroup included left lateralized parietal regions (Supramarginal Gyrus, Inferior Parietal Lobule). Notably, this included regions that have been argued to be involved in mapping from letters to sounds. The second subgroup also showed activation in temporal regions where the first subgroup did not, specifically the superior temporal pole, which has been argued to be involved in semantic storage, and bilateral frontal regions (inferior frontal gyrus pars triangularis and opercularis, insula, precentral gyrus), some of which have been implicated in semantic access and others which have been implicated in the articulatory aspect of spoken production. Finally, the second subgroup had large β-values in a number of subcortical regions (Thalamus, Pallidum, Caudate Nucleus, Midcingulate Gyrus) where the first subgroup did not. Overall, there were significant differences in accuracy for the scanner task between Group 1 and Group 2. Group 1 made significantly more errors (3.6%) than Group 2 (0.8%, $t(28) = 2.56$, $p = .016$).

Between-Group Univariate Analysis. The second fMRI analysis was a whole brain comparison looking at how an individual's response to word vs. pseudoword stimuli. First, we looked at an analysis that collapsed across subgroups, to ensure that the

findings of the current experiment matched previous studies. With a primary $p < .001$ and a Gaussian Random Field theory cluster-size correction to a family-wise error rate (FWE) of $p < .05$, we found 10 clusters that showed greater activation for words than for nonwords and 12 clusters that showed greater activation for nonwords than for words. Figures and tables with the details of the the results of these analyses are in the supplementary materials.

Briefly, for the words greater than nonwords contrast, we found a large cluster (>500 voxels) in the left angular gyrus and middle occipital temporal cortex, consistent with what has been previously reported in the literature (Binder et al., 2009). We also found moderately sized cluster (100-500 voxels) in the primary visual cortex, a cluster that is largely contained within the anterior portions of the left fusiform gyrus, and a right middle temporal gyrus cluster. There was also a moderately sized cluster in a medial parietal region in the left precuneus, with a smaller cluster (<100 voxels) in the homologous right hemisphere region. Finally, smaller clusters were found in the right supramarginal gyrus and three smaller right hemisphere clusters in the visual cortex (calcarine, lingual gyrus and superior occipital gyrus).

For the nonwords greater the word contrast, there were extremely large clusters in the both left (2706 voxels) and right (1253 voxels) frontal cortex, centering on the inferior frontal gyrus and the anterior insula cortex. There was also a large bilateral cluster (856 voxels) in the left and right medial aspect of the superior frontal gyrus. Two sizeable clusters were found in the left inferior parietal lobule, consistent with the findings of the Taylor et al. (2013) meta-analysis. The remaining clusters included two smaller right orbital frontal clusters, a larger right cerebellar cluster, a larger cluster in

the right angular gyrus and several subcortical clusters, including the left caudate and the right thalamus.

Critical for the current study, we next investigate how these patterns of activation varied as a function of subgroup assignment. With a primary $p < .005$ and a Gaussian Random Field theory cluster-size correction to a family-wise error rate (FWE) of $p < .05$, or a cluster extent threshold of k = 95, four significant clusters emerged showing this interaction, shown in **Figure 2** and in reported in Table 4. The largest cluster includes 735 voxels bilaterally in the occipital cortex, superior to the calcarine fissure and extended up in the cuneus gyrus. Looking at the peak voxel in this cluster, individuals in Subgroup 1 show a greater activation for words than for pseudowords, while individuals in Subgroup 2 show a greater activation for pseudowords than for words. Another cluster of 339 voxels falls largely in the right precuneus and right angular gryus. Like the primary visual area cluster, at the peak voxel in this cluster, Subgroup 1 shows greater activation for words than for pseudowords, while Subgroup 2 shows greater activation for pseudowords than for words. A third cluster of 132 voxels falls primarily in the left inferior parietal lobule, though extends slightly in the superior parietal lobule. Unlike the first two clusters, in this region both Subgroup 1 and Subgroup 2 show greater activation to pseudowords than to words, but Subgroup 2 shows a much larger effect. Critically, this inferior parietal lobule region aligns very closely with the region identified by the Taylor et al. (2013) meta-analysis as the region most likely to be involved in sublexical processing. Finally, a cluster of 115 voxels falls near the canonical visual word form area (VWFA) at the junction of the left inferior occipital and temporal cortex. As with the inferior parietal lobule cluster, in the VWFA cluster, subgroup 2 showed

greater activation for pseudowords than for words, while subgroup 1 showed equivalent activation for both types of orthographic stimuli. The Taylor et al. (2013) meta-analysis also identified this region as showing greater activation for pseudowords than for words. There were no regions in which Group 1 showed a greater difference between pseudowords and words than Group 2.

Representational Similarity Analysis. A final fMRI analysis used representational similarity analysis based on the 20 familiar words used in the experiment to map the phonological and semantic processing of different anatomical regions of interest, defined based on the Taylor et al. (2013) meta-analysis. Specifically, we looked at semantic and phonological processing in the left angular gyrus (AG) and the left inferior parietal lobule (IPL), the left inferior frontal gyrus (IFG) and the ventral occipitotemporal cortex (vOT). The result of the analysis that looks at all participants, ignoring subgroup membership, is shown in **Figure 3**. In all four of the ROIs, the correlation between the brain-based similarity matrix and the predicted phonological similarity matrix was statistically significantly greater than 0, using a one-tailed t-test (AG = .033, $t(29) = 1.83$, $p = .039$; IPL = .034, $t(29) = 1.72$, $p = .048$; IFG = .049, $t(29) = 2.20$, $p = .018$; vOT = .062, $t(29) = 2.45$, $p = .010$). The correlation between the group-average brain-based similarity matrix and the predicted semantic similarity matrix was statistically significant in the vOT region (.048, $t(29) = 1.76$, $p = .045$) and in the angular gyrus region (.03, $t(29) = 1.76$, $p = .036$). These results show that, at least in the context of a reading aloud task, there is clear phonological processing of written words in frontal speech production areas, the angular gyrus, the inferior parietal lobule and the vOT, and semantic processing in the vOT and the angular gyrus.

One particularly surprising aspect of these results is that the pattern of activation in the angular gyrus is correlated with both phonological and semantic processing. The angular gyrus has been identified as a part of the brain's semantic network (e.g. Binder et al., 2009), largely due to the fact that it shows greater activation to words than to pseudowords, as we show in our own results above and as Taylor et al. (2013) find in their meta-analysis, and because its activity correlates with word frequency and to a lesser extent imageablity (Graves et al., 2009). However, as Taylor and colleagues (2013) point out, these results are consistent with either a region that processes lexical phonological information or a semantic region, the current results suggest that this region is engaged with both of these functions during in reading aloud, though previous results from our lab (Fischer-Baum et al., 2017) reported orthographic processing in this region during a different reading task. An alternative interpretation is that there is variation across anatomical atlases for what is considered part of the angular gyrus (Seghier, 2013); perhaps in the broader inferior parietal lobule there are regions that are doing semantic processing of words and other regions that are doing phonological processing, and the precise boundaries between these regions do not fall clearly on anatomical divisions. The region that the Automatic Anatomical Labeling (AAL) brain atlas identifies as being the angular gyrus may, for example, include what other atlases might identify as supramarginal gyrus, which has been identified as playing a phonological role in reading (e.g. Stoeckel, Gough, Watkins & Devlin, 2009). The finding that our angular gyrus region is both phonological and semantic may have more to do with an inconsistent use of brain labels in the literature, rather than a finding that is inconsistent with previous claims.

Critical for current study, though, is whether there are differences in these RSA results as a function of subgroup assignment from the clustering analysis. For each ROI, subgroups were compared on the phonological and semantic similarity measures. Subgroup 2 had a significantly higher phonological similarity index in all four regions of interest (AG: -.02 vs. .06, t(28) = 2.24, *p* = .033; IFG: -.01 vs. .08, t(28) = 2.11, *p* = .049; IPL: -.03 vs. .07, t(28) = 2.46, *p* = .020; vOT: -.02 vs. .10, t(28) = 2.51, *p* = .018) Strikingly, the vOT did not show between group differences in semantics (Subgroup 1 = .02, Subgroup 2 = .06, , t(28) = .80 , *p* = .43), and neither did any of the other regions of interest (AG: -.01 vs. .05, t(28) = 1.93, *p* = .063; IFG: -.01 vs. .04, t(28) = .85, *p* = .404; IPL: -.02 vs. .04, t(28) = 1.45, *p* = .158).

Subgroups identified by the clustering analysis differ with respect to phonological processing, with Subgroup 2 showing higher correlations with the phonological similarity matrix than Subgroup 1 in all four regions of interest. However, the two groups do not differ in semantic processing of written words. It is worth noting that Subgroup 1 generally performed worse on the task in the scanner, meaning that they were less accurate at generating the correct phonological form, consistent with the RSA results.

Behavioral results by fMRI subgrouping. Finally, we reanalyzed the behavioral data based on the groupings generated from the fMRI data. We were primarily interested in between-group differences in the pseudohomophone, lexicality and regularity effects but we also tested whether the groups differed on standardized measures of reading skill.

The pseudohomophone experiment was reanalyzed using subgroup-membership as a between-subject factor. The results of this analysis are shown in **Figure 4**. Figure 4a shows reaction time for correct trials in the pseudohomophone lexical decision task,

for both pseudohomophones and other nonwords, by subgroup. There were no significant differences between subgroups 1 and 2 in the overall lexical decision time (Subgroup 1 = 587ms, Subgroup 2 = 606ms, t(27) = .77, *p* = .45). However, the group by nonword-type interaction was significant (F(1,26) = 5.01, *p* = .034). Specifically, individuals in Subgroup 1 had a robust difference between pseudohomophones and other nonwords (599 vs 577, t(9) = 4.98, *p* =.0008), while individuals in Subgroup 2 showed a small, marginally significant (617 vs 608, t(19) = 2.15, *p* = .060). These differences can be seen at the individual subject level. Figure 4b plots individual subject data of the size of the pseudohomophone effect for the two clusters, with median values indicated by a black line.

The reading aloud experiment was also reanalyzed using cluster-membership as a between-subject factor. The results of this analysis are shown in **Figure 5**. Figure 5a shows reaction time for regular words, irregular words and pseudoword reading time. There was a significant difference between groups 1 and 2 in reading time (Subgroup 1 = 493ms, Subgroup 2 = 554ms, t(27) = 2.46, *p* = .020), with subgroup 1 being faster will all types of stimuli (Regular Words = 32 ms; Irregular Words = 55 ms; Pseudowords = 97 ms). Still, the group by stimulus-type interaction was significant (F(2,54) = 4.94, *p* = .011). Individuals in Subgroup 2 showed a larger lexicality effect (all words vs. pseudowords: 106ms) than individuals in Subgroup 1 (53ms, t(27) = 2.40, *p* = .023), and also a trend in the direction of a larger regularity effect (irregular words vs. regular words; Subgroup 1: 8ms, Subgroup 2: 31ms, t(27) = 1.71, *p* = .099). Again the lexicality differences can be seen at the individual subject level, with Figure 5b plots individual subject data of the size of the lexicality effect for the two clusters, with median values

indicated by a black line. Figure 5c plots the same analysis for the regularity effect. For the simple main effects, Subgroup 2 was significantly slower than Subgroup 1 for exception words (535 ms vs. 479 ms, t(27) = 2.46, p = .021) and nonwords (625 ms vs. 529 ms, t(27) = 2.77, *p* = .010). While there was a trend towards being slower in regular word reaction time as well, the difference was not significant (503 ms vs. 471 ms, t(27) = 1.52, *p* = .14).

Finally, Table 5 reports between group differences on standardized reading measures. Subjects from the different subgroups did not differ significantly on any of the standardized reading measures, indicating that these other behavioral and neuroimaging differences did not correspond to differences in reading skill. One particularly surprising result from this analysis is the failure to find group differences on the PDE task, which is typically used to assess sublexical reading ability. One would expect scores on this task to be positively correlated with the other indices of sublexical reading, like the pseudohomophone effect and the regularity effect. As can be seen in Table 3, there is no correlation. One might also expect the subgroups to differ on the PDE task, if they differ on sublexical efficiency, but as can be seen in Table 5 there are no differences here either. One possible explanation for this is that nearly all of our participants performed near ceiling on this task. The phonemic decoding efficiency task might not be as sensitive to subtle differences in sublexical processing between highly literate readers as the other indices discussed above.

Discussion

Even within a range restricted set of highly competent, adult, literate readers, individuals vary in their performance on reading tasks. This variability was observed in high-level

measures of reading skill like the ability to comprehend passages and in more specific, cognitively informed measures, like the time it takes them to reject pseudowords that are pronounced like real words in a lexical decision task or the time it takes to read nonwords. Furthermore, these differences in reading behavior are linked to differences in how the brain responds to written words. Using a novel, data-driven approach to analyzing multi-subject, task-based fMRI data, two subgroups of readers were identified on the basis of how their brains response to words and pseudowords in a reading aloud task.

Follow up neuroimaging analyses with an orthogonal set of data suggest that these subgroups differed in phonological processing of written stimuli. First, the size of the pseudowords > words contrast was significantly larger for the second subgroup than the first subgroup in several regions of the brain, most notably in the left inferior parietal lobule and the left ventral occipitotemporal cortex. Second, the extent to which the similarity relationships among the fine-grained patterns of activation to specific words matched the phonological similarity between the words was greater in the second subgroup than the first subgroup in the left inferior frontal gyrus, the left angular gyrus, and the left inferior parietal lobule. Finally, the first group showed a larger pseudohomophone effect, that is they were slower at rejecting pseudowords that sound like real words (e.g., BRANE) than rejecting matched pseudowords (e.g. BRAME), than the second group, while the second group showed a larger lexicality effect, that is they were slower at reading pseudoword (e.g. PENK) than word stimuli (e.g. YARN), than the first group. Taken together, one interpretation of this finding is that participants in the

first cluster have more efficient and less effortful sublexical processing of words than individuals in the second cluster.

Our findings are particularly noteworthy in that the data used to cluster participants was fully independent of the data used to validate that clustering. The behavioral data that was collected to validate the fMRI clustering was collected during a different testing session, on average 2 weeks from the fMRI session, and included a variety of reading tasks, both reading aloud and lexical decision. As discussed above, previous research has shown that the reading system changes strategically as a function of task demands. Based on this prior research, we assume that participants in the current experiment are varying their lexical and sublexical reading routes based on task-demands during both the imaging task in the scanner and the behavioral tasks outside of the scanner. Despite that, the clustering algorithm applied to the neuroimaging task is picking up on some individual variability that appears to be stable across time and task.

What then is this stable difference between readers that is being identified in these experiments? Both behavioral and neuroimaging analyses suggest that the clustering algorithm has identified readers who differ in their sublexical reading route. Behaviorally, the group that has the larger lexicality effect also shows a significantly smaller pseudohomophone effect. The pseudohomophone effect has largely been interpreted to reflect the fact that when we are presented with pronounceable written stimuli, we automatically generate a phonological representation, even during tasks that do not require overt pronunciation like visual word recognition, which in turn influence other processes associated with word reading, like the activation of semantics (Frost,

1998; Van Orden, 1991; Braun et al., 2009; Leinenger, 2014; Harm & Seidenberg, 2004)[2]. This phonological activation poses a particular challenge in lexical decision which requires only knowledge of word spelling, with many theories proposing an additional "spelling check" step by which the brain compares a spelling predicted by activated phonology and semantics with the actual stimuli. Groups could differ in the size of the pseudohomophone effect either because they differ in the amount that phonological information is competing with orthographic information when making the lexical decision or they could differ in the "spelling check" mechanism. As we discuss below, both of these group differences could contribute to the observed findings.

One possibility is that readers in Subgroup 1 have a greater degree of competition between orthographic and phonological representations than readers in Subgroup 2. The degree of competition could be driven, in part, by the speed of the sublexical reading route relative to the lexical route. Subgroup 1 is generally faster at reading aloud than Subgroup 2, but the difference is more pronounced for nonwords (97ms difference) than exception words (55ms difference) or regular words (37ms difference). The readers who show a smaller pseudohomophone effects are also more delayed at generating pronunciations from pseudoword stimuli relative to the speed with which they generate pronunciations for word stimuli, perhaps because the use of the sublexical route is more effortful for this group. As a result, information from those pronunciations is less likely to interfere in the lexical decision task.

---

[2] In addition to this phonological explanation, Harm and Seidenberg (2004) suggest a limited contribution of direct mappings between orthography and semantics in the pseudohomophone effect. With appropriately matched nonword controls, as in the current study, these effects are very small and therefore we think that it is unlikely that these between group differences have to do with between group differences in orthography to semantics mapping.

Some of the results of the neuroimaging analyses also support this difference in processing between the two groups. Both groups show greater activation for pseudowords than words in a large cluster in the inferior parietal lobule, a region identified as being critical for sublexical processing during reading aloud (Taylor et al., 2013), but Subgroup 2 shows a greater difference in activation between the two types of stimuli than Subgroup 1. One way to interpret this difference is to assume that the groups differ in the effort that it takes to use the sublexical route. Taylor et al. (2013) propose a u-shaped BOLD response function on the basis of engagement and effort, with a lack of engagement showing the smallest BOLD response, effortful engagement showing the largest BOLD response, a greater response than less effortful engagement. If we assume that this region is engaged in the mapping from letters to sounds in both groups, then the difference between Subgroup 2 and Subgroup 1 is that this process for the same set of pseudowords is more effortful for Subgroup 2, which is consistent with the fact that Subgroup 2 is slower at pseudoword reading than Subgroup 1.

Subgroup 2 also shows more pseudoword than word activation that Subgroup 1 in other cortical regions, for example in the visual word form area, a region that has been postulated to represent sublexical orthographic form (Dehaene et al., 2005; Fischer-Baum et al., 2017). In reading aloud studies, it is typical for this region to show greater activation for pseudowords than for words (Taylor et al., 2013), though equivalent word and pseudoword activity is often reported in passive tasks that do not require overt naming (e.g. Cohen et al., 2002). Taylor and colleagues (2014) interpreted this difference as reflecting top-down signals from regions that are generating phonological information from unfamiliar words (see also Price & Devlin, 2011). These

proposed top-down signals operate like the “spelling check” function described above; that is, the top-down signal is generated the prediction of the word’s spelling and there is greater activation when there is a mismatch between the stimuli and the prediction. Under this interpretation, individuals in Subgroup 2 are using this top-down signal more than individuals in Subgroup 1. Note that this result is in the context of a reading aloud experiment without pseudohomophones. Therefore, the between group difference in the pseudohomophone effect could be explained by a difference in the reliance on the top-down “spelling check” signal.

In addition to looking for differences in univariate activation, we looked for between group differences in information processing using representational similarity analysis. There were no differences between the groups in the degree of semantic information elicited by the written words. The groups did show differences in phonological information processing in the left vOT, inferior parietal lobule, the angular gyrus, and the inferior frontal gyrus. Specifically, Subgroup 2, which showed both behavioral and univariate neuroimaging evidence for more effortful sublexical processing, showed higher correlations between the brain-based similarity structure in those regions and the predicted similarity structure based on phonology. On the surface, this result may appear to go in the opposite direction of what would be predicted by the group for whom sublexical processing is less effortful. However, given that the task in the scanner is a reading aloud task, it is clear that both groups are activating phonological representations of the written words. Some of the words being read aloud in the scanner had irregular spelling-to-sound correspondences (e.g. YACHT). The phonological similarity matrix is based only a word’s received

pronunciation (e.g., /jɑt/), which is generated by the lexical reading route. For these irregular words, the sublexical reading route may be activating alternative pronunciations of the written word (e.g. /jækt/, /jætʃt/, /jæt/, etc.) which compete with the received pronunciation. Readers in Subgroup 2, whose sublexical processing requires more effort, will be activating the received pronunciation more quickly than these alternative pronunciations. As a result, for Subgroup 2, the phonological representation in these regions is closer to the received pronunciation and therefore correlates higher with the phonological similarity matrix. For Subgroup 1, the phonological representation in these regions will be farther from the received pronunciation and therefore correlate less with the phonological similarity matrix. This interpretation is particularly compelling if we assume that the angular gyrus plays a role in processing lexical phonological information (Taylor et al., 2013), and therefore its activation reflects selection of the received pronunciation rather than the generation of a possible pronunciation by the sublexical route. While not specifically designed to contrast regular and irregular words, it is worth noting that many of the word stimuli in the in-scanner task were irregular (e.g. YACHT, EYE, SEW). An additional prediction of this account would be that differences in phonological similarity between Subgroup 1 and Subgroup 2 in these regions would be driven by the irregular words only. However, the current experiment is underpowered to examine this prediction.

This interpretation is challenged a little by the observation that readers in the two subgroup differ in ways beyond our indices of sublexical processing. In the scanner, readers in Subgroup 1 performed more poorly, and outside of the scanner, they were significantly faster overall in the reading aloud study and numerically faster in the lexical

decision study. Therefore, these subgroups may differ on overall processing speed or perhaps on where they place the criterion in a speed/accuracy tradeoff. Faster reading time could explain the fact that the difference between the two groups identified by the clustering algorithm was one of degree rather than dissociation. Longer reaction times might result in larger BOLD responses, meaning that the algorithm might have clustered on the basis of faster versus slower readers rather than differences in reading strategies (Binder et al., 2005). While this difference may explain some of our results, we think that this interpretation is unlikely the only account for all of our findings. The standardized reading batteries should be sensitive to differences in overall processing speed or in speed/accuracy tradeoff, but no differences were observed. Furthermore, it is unclear how this account could explain the behavioral interactions, with Subgroup 1 showing a particularly large pseudohomophone effect in reaction time and Subgroup 2 showing even greater slowing for pseudowords compared to regular word stimuli.

We conclude that our current study suggests individual differences in how effortful engagement of the sublexical pathway is in highly literate adult readers. We can only speculate as to the source of this variability. Woollams and colleagues (2016) suggest that developing readers come into the learning process with differences in phonological processing abilities. Children with stronger phonological representations are more likely to learn to read through a sublexical dominant pathway, while children with weaker phonological representations will tend to rely on the lexical pathway. These differences in the initial stages of learning are still reflected in how words are processed in the adult brain. While we did not collect any measures of phonological processing ability in the current study, Luque, Luque and Lopez-Zamora (2011) showed that the

size of pseudohomophone effect varied with individual's categorical perception ability, with those individuals with worse phonological processing ability also showing a reduced pseudohomophone effect (see also Holyk and Pexman, 2004). This finding is consistent with the proposal laid out above. If we assume that adults with more effortful phonological processing from written words were also children with weaker phonological processing ability, then learning to read with weaker phonological representations results in a slower, more effortful sublexical/phonological route, which in turn diminishes the pseudohomophone effect.

Strikingly, this difference between readers does not appear to map onto any differences in reading skill, suggesting that there are multiple reading modes that can be equally effective organizations of the reading network. These findings run contrary to claims of the stage-based theory of reading development that assumes that young readers first develop their sublexical route, and later develop more the efficient lexical route (Share, 1995; Ehri, 1992). Indeed, as children learn to read better the size of the pseudohomophone effect is reduced (Grainger et al., 2012; Sprenger-Charolles et al., 2003) indicating less of a reliance on the sublexical/phonological route. Extending this claim to adult readers, we would predict that relying more on sublexical/phonological route, and therefore showing a greater pseudohomophone effect, would be related to worse overall reading skill. However, we find no differences between the two clusters – defined by fMRI but also differing on pseudohomophone and lexicality effect size – in terms of standard measures of reading performance in either the TOWRE-2 or the Nelson-Denny Test.

Taken together, our results suggest that there is heterogeneity even within a group of highly-skilled readers. This heterogeneity is detectable in behavioral measures of reading ability and it is detectable in the brain's response to written words. Heterogeneity of this type has important implications for our understanding of reading in the brain. For example, much of the research into developmental dyslexia asks how reading disordered individuals, as a group, deviate from neurotypical readers, as a group, in both behavioral and neuroimaging measures (Ramus, 2003; Richlan, 2012). Serious concerns have been raised about treating individuals with dyslexia as a homogenous group (Hadzibeganovic et al., 2010) as a range of different underlying impairments, with corresponding differences in brain structure and function, could lead to reading disabilities. A similar concern could be raised for treating neurotypical readers as a homogenous group, as it is clear from the current study that individuals vary in how their brain responds to written words. Based on our current study it seems like there are two subtypes of readers. It is premature to draw conclusions about the precise prevalence of reading subtypes in the general population from the sample used in our experiment. However, it is unlikely that differences in reading style would be classified as what de Schotten and Shallice (2017) call a "minority-discrepant situation", that is a situation in which most individuals show one pattern with only a few outlying subjects. Instead, this difference appears to be a "major-discrepancy situation", with multiple subtypes of readers frequently observed in the population. As de Schotten and Shallice (2017) discuss, major discrepant situations have the largest consequences for drawing inferences from group-average data.

The individual differences in brain activation reported in the current study likely reflects only a fraction of the heterogeneity that exists in literate adults. Welcome and Joanisse (2012) also found that activation to other cortical regions (e.g. precuneus, middle temporal gyrus) correlate with a measure of lexical reading ability (SWE), while we found no differences between our subgroups on lexical measures. One limitation of the clustering approach taken here is that readers may differ on a number of dimensions. Our analysis may have picked up on the primary dimension that distinguishes readers in our sample – the effort needed to use the sublexical reading route – but readers in our sample may also differ on the effort needed to is the lexical pathway as well. Furthermore, activation patterns are not the only source of hetereogeneity in the reading brain that could be related to individual differences in reading behavior. Structural differences in white matter connectivity have also been linked to differences in sublexical reading ability (Welcome and Joanisse, 2014) and semantic processing when reading (Graves et al., 2014). In general, understanding the ways that adult, neurotypical readers vary both in their reading behavior and their reading brains provide clear benchmarks that will allow existing models of reading in the brain to extend to explaining individual differences.

Acknowledgements. Funding for this project was provided by NSF SES-1659925 (M.V.), NSF BCS- 1752751 (S.F.B.), NIH R21DC016715 (S.F.B), the T.L.L. Temple Foundation (S.F.B. and A.R.N) and the Social Science Research Institute of Rice University (M.V. and S.F.B.).

Author Contribution. S.F.B. and M.V designed the study and interpreted the results. S.F.B. wrote the initial draft of the manuscript, though all authors contributed to subsequent draft. J.H.K designed and implemented the novel fMRI statistical analysis. Y.L. performed the behavioral experiments and analyzed the data. A.R.-N. performed the neuroimaging experiments and preprocessed the neuroimaging data

References

Andrews S. 2012. Individual differences in skilled visual word recognition and reading. In: Adelman, JS editor. Visual Word Recognition, 2, New York: Psychology Press.

Baron J, Strawson C. 1976. Use of orthographic and word-specific knowledge in reading words aloud. J Exp Psychol Human Percept Perform 2: 386-393.

Besner D, Davelaar E. 1983. Suedohomofoan effects in visual word recognition: Evidence for phonological processing. Can J Psychology 37: 300-305.

Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., & Liebenthal, E. (2005). Some neurophysiological constraints on models of word naming. *Neuroimage, 27(3),* 677-693.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767-2796.

Bitan, T., Booth, J. R., Choy, J., Burman, D. D., Gitelman, D. R., & Mesulam, M. M. (2005). Shifts of effective connectivity within a language network during rhyming and spelling. *Journal of Neuroscience*, *25*(22), 5397-5403.

Braun M, Hutzler F, Ziegler JC, Dambacher M, Jacobs AM. 2009. Pseudohomophone effects provide evidence of early lexicophonological processing in visual word recognition. Hum Brain Mapp 30: 1977-1989.

Brown JI, Fishco VV, Hanna G. 1993. Nelson-Denny reading test: Manual for scoring and interpretation, forms G & H. Riverside Publishing Company.

Brown P, Lupker SJ, Colombo L. 1994. Interacting sources of information in word naming: A study of individual differences. J Exp Psychol Human Percept Perform 20: 537-554.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054-1069.

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. 2001. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol Rev 108: 204-256.

Coltheart M, Rastle K. 1994. Serial processing in reading aloud: Evidence for dual-route models of reading. J Exp Psychol Human Percept Perform 20: 1197-1211.

De Schotten MT, Shallice T. 2017. Identical, similar or different? is a single brain model sufficient?. Cortex 86: 172-175.

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.

Ehri LC. 1992. Reconceptualizing the development of sight word reading and its relationship to recoding. In: Gough PF, Ehri LC, Treiman, R editors. Reading Acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates

Fischer-Baum S, Bruggemann D, Gallego IF, Li DS, Tamez ER. 2017. Decoding levels of representation in reading: A representational similarity approach. Cortex 90: 88-102.

Forster KI, Forster JC. 2003. DMDX: A Windows display program with millisecond accuracy. Behav Res Methods 35: 116-124.

Frost R. 1998. Toward a strong phonological theory of visual word recognition: true issues and false trails. Psychol Bull 123: 71-99.

Gough PB, Tunmer WE. 1986. Decoding, reading, and reading disability. Rem Spec Ed 7: 6-10.

Grainger J, Lété B, Bertand D, Dufau S, Ziegler JC. 2012. Evidence for multiple routes in learning to read. Cognition 123: 280-292.

Graves WW, Binder JR, Desai RH, Humphries C, Stengel BC, Seidenberg MS. 2014. Anatomy is strategy: Skilled reading differences associated with structural connectivity differences in the reading network. Brain Lang 133: 1-13.

Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., & Binder, J. R. (2009). Neural systems for reading aloud: a multiparametric approach. *Cerebral Cortex*, *20*(8), 1799-1815.

Hadzibeganovic T, van den Noort M, Bosch P, Perc M, van Kralingen R, Mondt K, Coltheart M. 2010. Cross-linguistic neuroimaging and dyslexia: a critical view. Cortex 46: 1312-1316.

Hall KC, Allen B, Fry M, Mackie S, McAuliffe M. 2016. Phonological CorpusTools, Version 1.2. [Computer program].

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

Holyk GG, Pexman PM. 2004. The elusive nature of early phonological priming effects: Are there individual differences?. Brain Lang 90: 353-367.

Jobard G, Crivello F, Tzourio-Mazoyer N. 2003. Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. Neuroimage 20: 693-712.

Jobard G, Vigneau M, Simon G, Tzourio-Mazoyer N. 2011. The weight of skill: interindividual variability of reading related brain activation patterns in fluent readers. J Neurolinguist 24: 113-132.

Johnston TC, Kirby JR. 2006. The contribution of naming speed to the simple view of reading. Read Writ 19: 339-361.

Katz L, Brancazio L, Irwin J, Katz S, Magnuson J, Whalen DH. 2012. What lexical decision and naming tell us about reading. Read Writ, 25: 1259-1282.

Kherif F, Josse G, Seghier ML, Price CJ. 2009. The main sources of intersubject variability in neuronal activation for reading aloud. J Cognitive Neurosci 21: 654-668.

Kook J, Guindani M, Zhang L, Vannucci M. 2017. NPBayes-fMRI: Nonparametric Bayesian general linear models for single- and multi-subject fMRI data. Submitted for publication.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neurosci 12: 535-540.

Landauer TK, Foltz PW, Laham D. 1998. An introduction to latent semantic analysis. Discourse Process 25: 259-284.

Leinenger M. 2014. Phonological coding during reading. Psychol Bull 140: 1534-1555.

Lupker SJ, Brown P, Colombo L. 1997. Strategic control in a naming task: Changing routes or changing deadlines?. J Exp Psychol Learn Mem Cogn 23: 570-590.

Luque D, Luque JL, López-Zamora M. 2011. Individual differences in pseudohomophony effect relates to auditory categorical perception skills. Learn Individ Differ 21: 210-214.

Monsell S, Patterson KE, Graham A, Hughes CH, Milroy R. 1992. Lexical and Sublexical Translation of Spelling to Sound. J Exp Psychol Learn Mem Cogn 18: 452-467.

Perfetti C. 2007. Reading ability: Lexical quality to comprehension. Scientific Studies of Reading 11: 357-383.

Perfetti CA. 2011. Phonology is critical in reading. In: Brady SA, Braze D, Fowler CA, editors. Explaining individual differences in reading: Theory and evidence. New York: Psychology Press.

Plaut DC, McClelland JL, Seidenberg MS, Patterson K. 1996. Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychol Rev 103: 56-115

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

Ramus F. 2003. Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction?. Curr Opin Neurobiol 13: 212-218.

Rastle K, Coltheart M. 1999. Serial and Strategic Effects in Reading Aloud. J Exp Psychol Human Percept Perform 25: 482-503.

Rayner K, Foorman BR, Perfetti CA, Pesetsky, D, Seidenberg, MS. 2002. How should reading be taught?. Sci Am 286: 70-77.

Reich L, Szwed M, Cohen L, Amedi A. 2011. A ventral visual stream reading center independent of visual experience. Curr Biol 21: 363-368.

Richlan F. 2012. Developmental dyslexia: dysfunction of a left hemisphere reading network. Front Hum Neurosci 6: 120.

Rothlein D, Rapp B. 2014. The similarity structure of distributed neural responses reveals the multiple representations of letters. NeuroImage, 89: 331-344.

Rubenstein H, Lewis SS, Rubenstein MA. 1971. Evidence for phonemic recoding in visual word recognition. J Verb Learn Verb Be 10: 645-657.

Rueckl JG, Paz-Alonso PM, Molfese PJ, Kuo WJ, Bick A, Frost, SJ, Hancock R, Wu DH, Mencl WE, Duñabeitia JA and others. 2015. Universal brain signature of proficient reading: Evidence from four contrasting languages. Proc Natl Acad Sci U S A 112: 15510-15515.

Schlaggar BL, McCandliss BD. 2007. Development of neural systems for reading. Annu Rev Neurosci 30: 475-503.

Seghier ML, Lee HL, Schofield T, Ellis CL, Price CJ. (2008). Inter-subject variability in the use of two different neuronal networks for reading aloud familiar words. NeuroImage 42: 1226-1236.

Seghier, M. L., Kherif, F., Josse, G., & Price, C. J. (2011). Regional and hemispheric determinants of language laterality: implications for preoperative fMRI. *Human brain mapping*, *32*(10), 1602-1614.

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist, 19(1),* 43-61.

Share DL. 1995. Phonological recoding and self-teaching: Sine qua non of reading acquisition. Cognition 55: 151-218.

Sprenger-Charolles L, Siegel LS, Béchennec D, Serniclaes W. 2003. Development of phonological and orthographic processing in reading aloud, in silent reading, and in spelling: A four-year longitudinal study. J Exp Child Psychol 84: 194-217.

Stoeckel, C., Gough, P. M., Watkins, K. E., & Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition. *Cortex*, *45*(9), 1091-1096.

Taylor JS, Rastle K, Davis MH. 2013. Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. Psychol Bull 139: 766-791.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2014). Interpreting response time effects in functional imaging studies. *Neuroimage*, *99*, 419-433.

Torgesen JK, Wagner RK, Rashotte CA. 2012. TOWRE-2: Test of Word Reading Efficiency – Second Edition. Austin, TX: Pro-Ed.

Van Orden GC. 1991. Phonologic mediation is fundamental to reading. In: Besner D, Humphreys GW editors. Basic processes in reading: Visual word recognition. New York: Routledge.

Vul E, Pashler H. 2012. Voodoo and circularity errors. Neuroimage 62: 945-948.

Wagenmakers EJ, Ratcliff R, Gomez P, McKoon G. 2008. A diffusion model account of criterion shifts in the lexical decision task. J Mem Lang 58: 140-159.

Welcome SE, Joanisse MF. 2012. Individual differences in skilled adult readers reveal dissociable patterns of neural activity associated with component processes of reading. Brain Lang 120: 360-371.

Welcome SE, Joanisse MF. 2014. Individual differences in white matter anatomy predict dissociable components of reading skill in adults. NeuroImage 96: 261-275.

Woo CW, Krishnan A, Wager TD. 2014. Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations. Neuroimage 91: 412-419.

Woollams AM, Lambon Ralph MA, Madrid G, Patterson KE. 2016. Do you read how I read? Systematic individual differences in semantic reliance amongst normal readers. Front Psychol 7.

Yap MJ, Balota DA, Sibley DE, Ratcliff R. 2012. Individual differences in visual word recognition: insights from the English Lexicon Project. J Exp Psychol Human Percept Perform 38: 53-79.

Zevin JD, Balota DA. 2000. Priming and attentional control of lexical and sublexical pathways during naming. J Exp Psychol Learn Mem Cogn 26: 121-135.

Zhang L, Guindani M, Versace F, Engelmann JM, Vannucci M. 2016. A spatiotemporal nonparametric Bayesian model of multi-subject fMRI data. Ann Appl Stat 10: 638-666.

Zhang L, Guindani M, Versace F, Vannucci M. 2014. A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. NeuroImage 95: 162-175.

Zhao L, Chen C, Shao L, Wang Y, Xiao X, Chen C, Zevin J, Xue G. 2016. Orthographic and phonological representations in the fusiform cortex. Cerebl Cortex: 1-14.

Figure Legends:

Figure 1. Results of the nonparametric Bayesian model analysis treating all words as a single stimulus type. (a) Cluster dendrogram obtained with hierarchical cluster under the linkage method. Two clear clusters of subjects emerge with a single subject (10) appearing as an outlier. (b) Posterior group-level maps of β for the two largest clusters at 6 axial slices (top two rows) as well as the difference between the two clusters (bottom).

Figure 2. Sagittal slices of the left hemisphere regions that show a significant between-group interaction in the pseudoword vs. word comparison (primary $p < .005$, cluster-size threshold $p_{FWE} < .05$). Group is identified based on a clustering of participants on their BOLD responses to all word and pseudoword stimuli in the first run of the experiment. Pseudoword vs. word contrast carried out over runs 2-6.

Figure 3. Results of the representational similarity analysis. (a) Illustration of the analysis approach, calculating the brain-based correlation for each participant in each ROI, and then comparing the brain-based correlation with the semantic and phonological similarity measures (b) Regions of interest used in the RSA analysis. Red is left angular gyrus (AG), yellow is left inferior parietal lobule (IPL), blue is left inferior frontal gyrus (IFG) and green is the left ventral occipitotemporal cortex (vOT) (c) Results of the whole-group analysis for four regions of interest. Error bars reflect +/- 1 SEM.

Figure 4. Behavioral results in the pseudohomophone effect lexical decision experiment treating cluster assignment from the single stimulus type analysis as a between group variable. (a) Group differences in RTs for the pseudohomophone and other pseudoword

stimuli. Error bars are +/- 1 SEM. (b) Plot of the median pseudohomophone effect (black bar) and individual subject data for individuals in each cluster.

Figure 5. Behavioral results in the reading aloud regular, irregular and pseudowords experiment defining groups based on the clustering from fMRI data. (a) Group differences in RTs for the pseudohomophone and other pseudoword stimuli. Error bars reflect +/- 1 SEM. (b) Plot of the median lexicality effect (black bar) and individual subject data for individuals in each cluster. (c) Plot of the median regularity effect (black bar) and individual subject data for individuals in each cluster.

		Raw Score		Percentile	
		Mean	Range (IQR)	Mean	Range (IQR)
Nelson-Denny Form G	Reading Comprehension	35.1/38	29-38 (2.8)	89th	35-99 (14)
	Vocabulary Score	73.7/80	57-80 (5)	90th	51-99 (9.5)
	Reading Rate	317wpm	164-552 (146.3)	72nd	10-99 (36)
TOWRE-2	Sight Word Efficiency	103.1/108	80-108 (8)	85th	25-98 (21)
	Phonemic Decoding Efficiency	61.0/66	49-66 (4.8)	85th	47-98 (15.5)

Table 1: Descriptive statistics for standardized measures of reading skill.

		RT (SD)
Lexical Decision	Words	588 (61.2)
	Pseudohomophones	611 (74.1)
	Matched Pseudowords	598 (73.6)
Reading Aloud	Regular Words	492 (55.9)
	Irregular Words	515 (62.3)
	Pseudowords	591 (102.4)

Table 2: Mean reaction times (RT) in milliseconds with standard deviations in parentheses for the pseudohomophone lexical decision task and the reading aloud task.

	Reading Comp.	Vocab. Score	Reading Rate	SWE	PDE	Pseudo homophone	Lexicality	Regularity
Reading Comprehension	1.0							
Vocabulary Score	.33†	1.0						
Reading Rate	.26	.07	1.0					
Sight Word Efficiency (SWE)	.03	-.22	.42*	1.0				
Phonemic Decoding Efficiency (PDE)	.59***	.03	.50**	.39*	1.0			
Pseudohomophone Effect	.27	.15	.12	.21	.25	1.0		
Lexicality	.17	.07	.21	.02	.33†	.06	1.0	
Regularity	.00	.03	-.19	-.07	.08	.04	.51**	1.0

$^{\dagger}p < .1$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

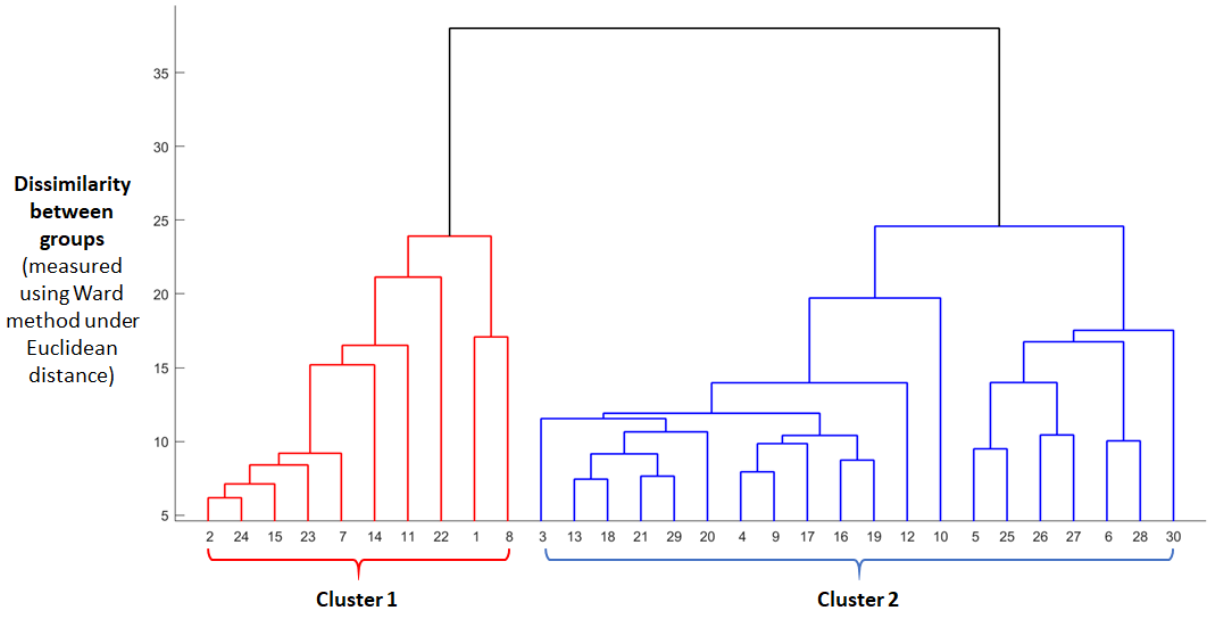
Table 3: Correlations between different behavioral reading measures

Clusters	Cluster Size	MNI Coordiates [x, y, z]	Subgroup 1	Subgroup 2
Right Precuneus/ Angular Gyrus	735	[16, -64, 40] *[32, -56, 40]* *[38, -60, 50]*	Words > Pseudo	Pseudo > Words
Bilateral Primary Visual Cortex	303	[-12, -82, 13] *[-24, -74, 14]* *[14, -68, 16]*	Words > Pseudo	Pseudo > Words
Left Inferior Parietal Lobule	132	[-32, -52, 48] *[-36, -58, 58]*	Words = Pseudo	Pseudo >> Words
Visual word form area	115	[-46, -60, -12] *[-52, -60, -20]* *[-60, -60, 18]*	Words = Pseudo	Pseudo >> Words

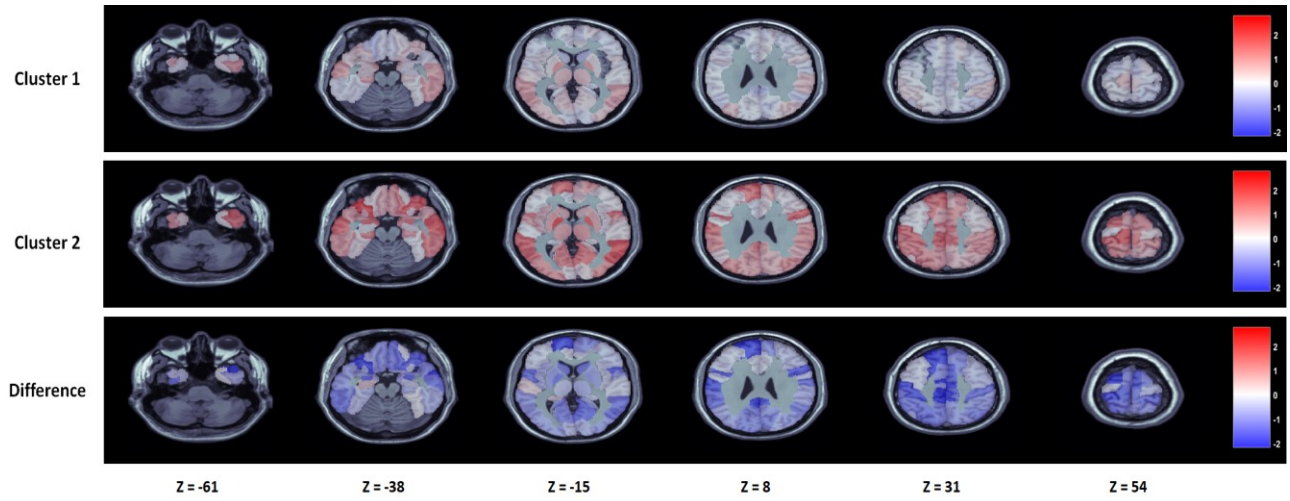
Table 4: Significant clusters of between-group interaction of pseudoword vs. word contrasts (primary $p < .005$, Gaussian Random Field cluster-size threshold $p_{FWE} < .05$), including the MNI coordinate of the peak voxel and at most two other local maxima (> 8mm distance). Relation of word to pseudoword activation for each subgroup at the peak voxel is also reported.

	Cluster 1 (n= 10)	Cluster 2 (n = 20)
Reading Comprehension	93.7%	91.6%
Vocabulary Score	94.0%	91.2%
Reading Rate	317	317
Sight Word Efficiency	101.5	103.9
Phonemic Decoding Efficiency	61.3	60.8

Table 6: Differences in standardized reading scores by group.


Dissimilarity between groups (measured using Ward method under Euclidean distance)
35
30
25
20
15
10
5
2 24 15 23 7 14 11 22 1 8
3 13 18 21 29 20 4 9 17 16 19 12 10 5 25 26 27 6 28 30
Cluster 1
Cluster 2

(a)


Cluster 1
Cluster 2
Difference
Z = -61
Z = -38
Z = -15
Z = 8
Z = 31
Z = 54

(b)

Figure 1

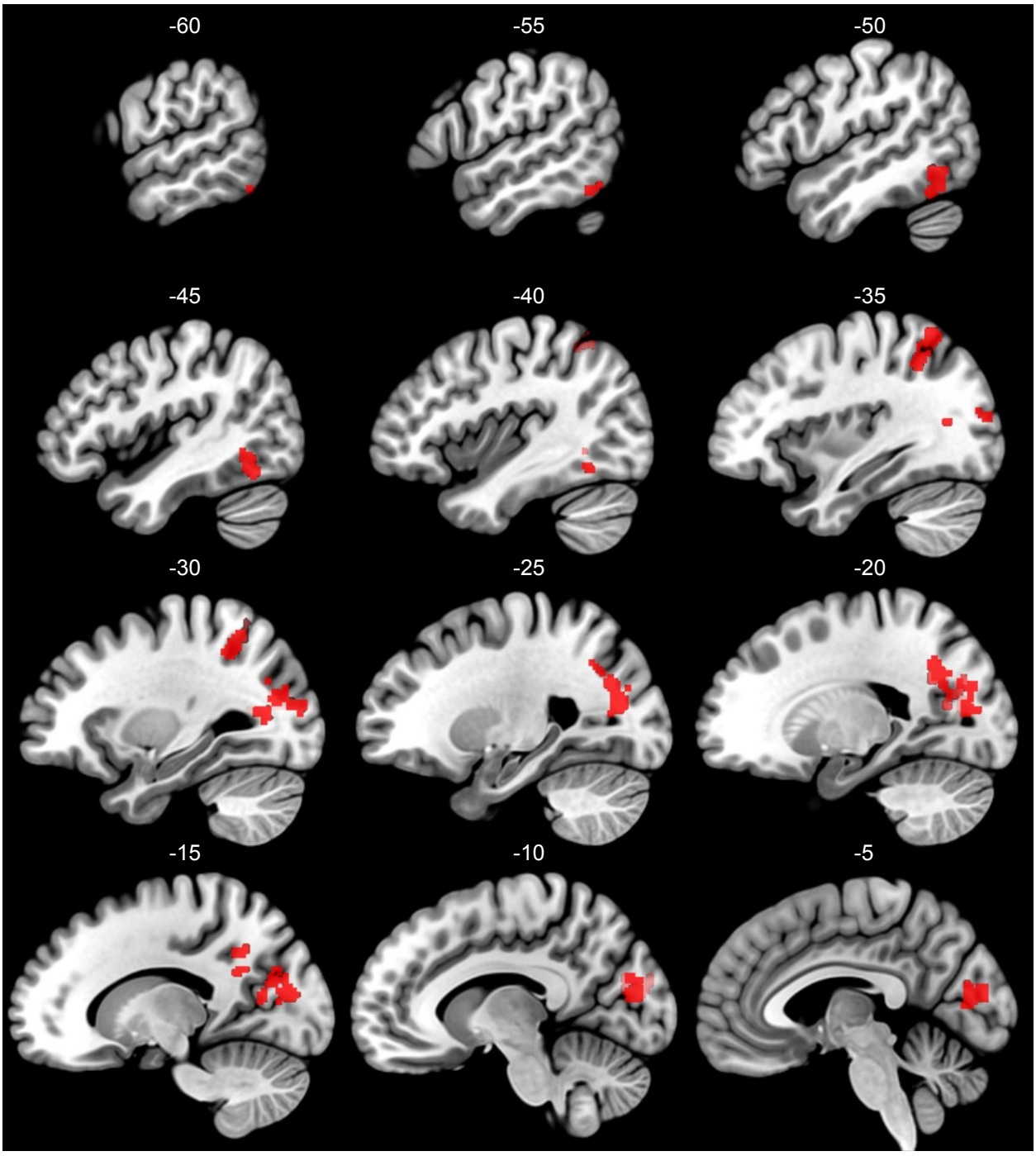
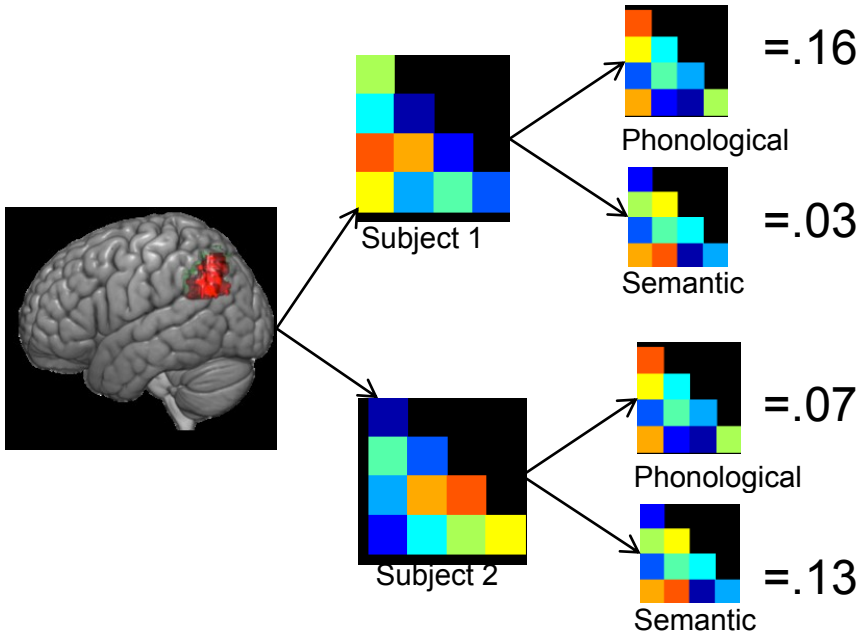
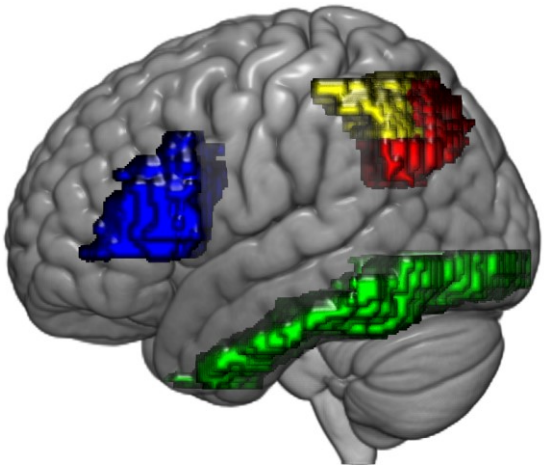


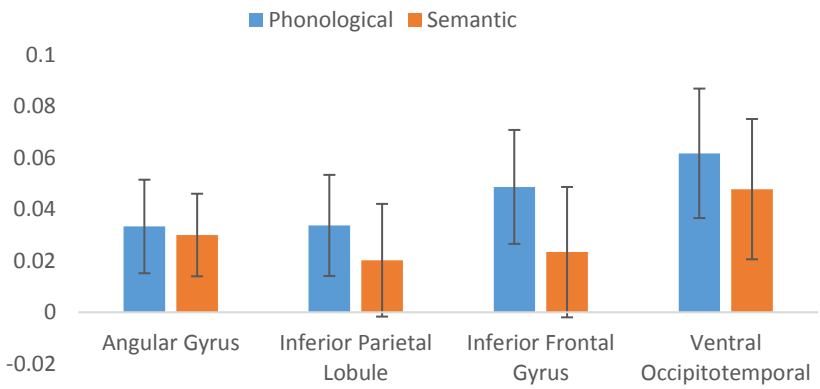
Figure 2


=.16
Phonological
Subject 1
=.03
Semantic
=.07
Phonological
Subject 2
=.13
Semantic

(a)

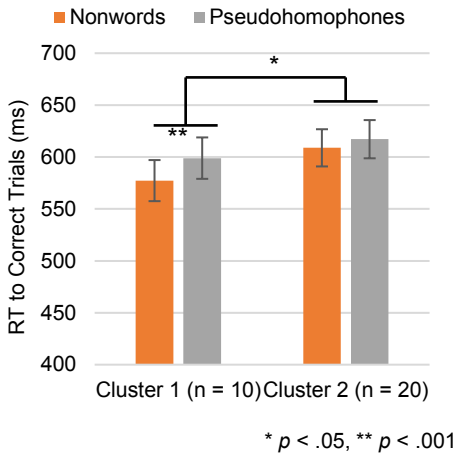


(b)

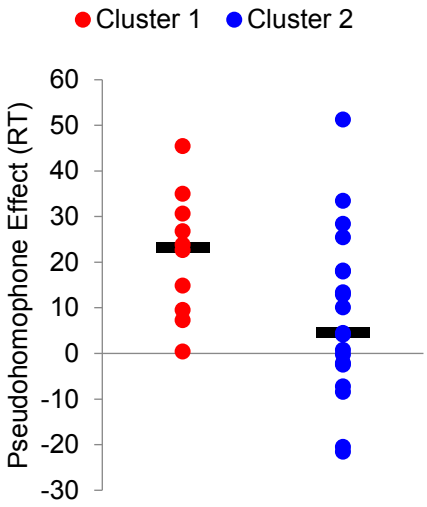

Phonological
Semantic
0.1
0.08
0.06
0.04
0.02
0
-0.02
Angular Gyrus
Inferior Parietal Lobule
Inferior Frontal Gyrus
Ventral Occipitotemporal

(c)

Figure 3

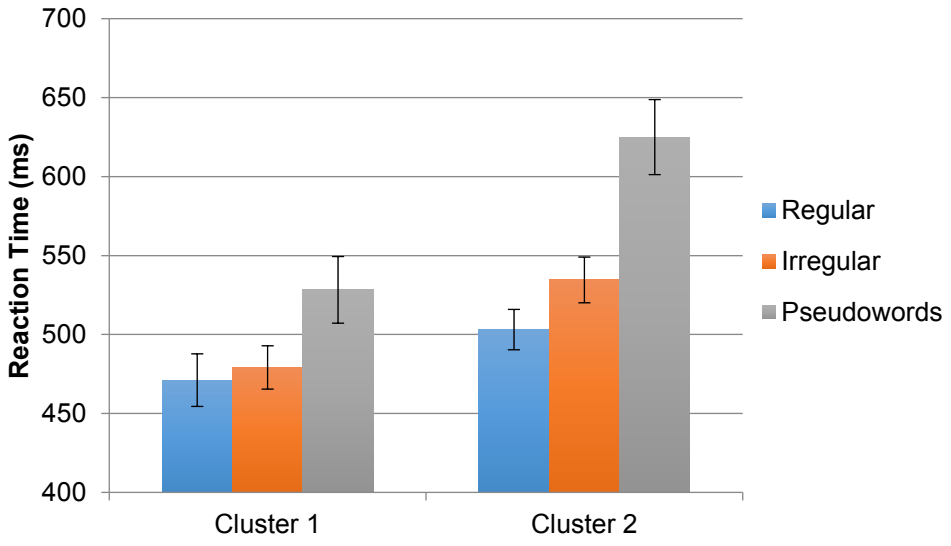

Nonwords
Pseudohomophones
700
650
600
550
500
450
400
RT to Correct Trials (ms)
*
**
Cluster 1 (n = 10)
Cluster 2 (n = 20)
* p < .05, ** p < .001

(a)

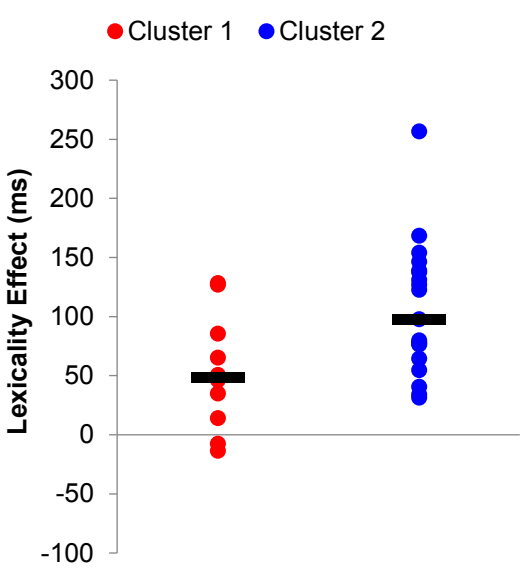

Cluster 1
Cluster 2
60
50
40
30
20
10
0
-10
-20
-30
Pseudohomophone Effect (RT)

(b)

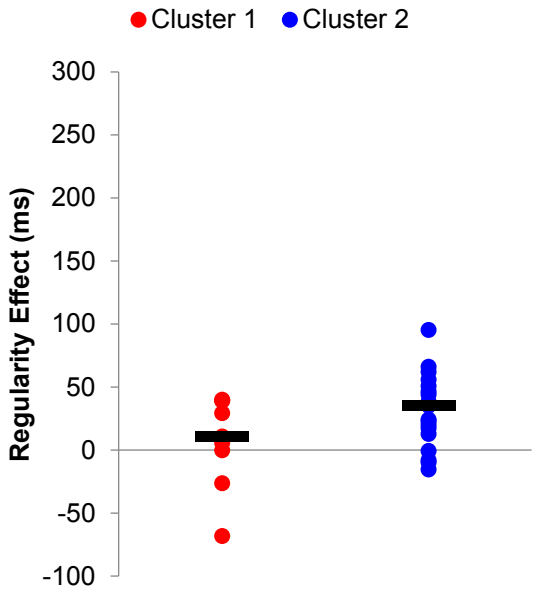
Figure 4


700
650
600
550
500
450
400
Reaction Time (ms)
Cluster 1
Cluster 2
Regular
Irregular
Pseudowords

(a)


Cluster 1
Cluster 2
300
250
200
150
100
50
0
-50
-100
Lexicality Effect (ms)

(b)


Cluster 1
Cluster 2
300
250
200
150
100
50
0
-50
-100
Regularity Effect (ms)

(c)

Figure 5

Appendix A

Stimuli used in fMRI experiment

Words		Nonwords
BLUE	KNEE	TEG
BOOT	LAUGH	SUT
CAPE	LYNX	STET
CLAM	PEACH	SITH
CREW	PEAR	ROIN
EYE	PIE	NINT
FLY	SEW	MULP
FOOT	SHOE	GEAD
FRUIT	SIGHT	DOIL
GRAY	YACHT	CHUT

Supplemental Materials

Clusters	Cluster Size	MNI Coordiates [x, y, z]
Left Angular Gyrus/Middle Occipital Cortex	937	[-42, -72, 22] *[-42, -76, 30]* *[-40, -60, 28]*
Primary Visual Cortex	371	[-14, -52, 16] *[-12, -52, 8]* *[-6, -62, 14]*
Precuneus	188	[0, -54, 10] *[-4, -60, 20]* *[-6, -50, -16]*
Right middle temporal gyrus	144	[50, -66, 10] *[46, -68, 24]* *[50, -58, 18]*
Left parahippocampal/fusiform gyrus	132	[-32, -42, -8] *[-28, -48, -4]* *[-30, -54, -10]*
Right calcarine	93	[8 -54, 10] *[10, -60, 20]* *[22, -58, 16]*
Right lingual gyrus	91	[20, -74, -6] *[18, -82, -8]* *[10, -86, -6]*
Right superior temporal/supramarginal gyrus	88	[60, -32, 24] *[54, -36, 28]*
Right superior occipital gyrus	67	[18, -86, 20]
Right precuneus	50	[8, -58, 58]

Table S1: Significant clusters of word > pseudoword contrast collapsing across all subjects (primary $p < .001$, Gaussian Random Field cluster-size threshold $p_{FWE} < .05$), including the MNI coordinate of the peak voxel and at most two other local maxima (> 8mm distance).

Clusters	Cluster Size	MNI Coordiates [x, y, z]
Left inferior frontal gyrus/Insula	2706	[-34, 16, 12] *[-56, 12, 30]* *[-38, 4, 26]*
Right inferior frontal gyrus/Insula	1253	[-34, 16, 12] *[-56, 12, 30]* *[-38, 4, 26]*
Superior frontal gyrus (medial)	856	[-4, 20, 42] *[-8, 8, 60]* *[-2, 4, 66]*
Left inferior parietal lobule	364	[-50, -40, 58] *[-56, -32, 50]* *[-64, -18, 32]*
Midcingulate area	250	[0, -14, 28] *[4, -34, 26]* *[0, -24, 32]*
Right Cerebellum (lobule VI and crus I)	179	[26, -68, -24] *[42, -68, -24]* *[20, -72, -18]*
Right angular gyrus	168	[36, -68, 50] *[36, -62, 56]* *[26, -62, 46]*
Left inferior parietal lobule	113	[-28, -46, 42] *[-34, -50, 48]* *[-34, -52, 40]*
Right Thalamus	79	[8, -6, 6] *[10, -10, 14]* *[10, -8, -2]*
Right middle orbital frontal gyrus	67	[30, 62, -2] *[36, 48, -10]* *[40, 60, 0]*
Right superior orbital frontal gyrus	64	[24, 40, -12] *[24, 50, -10]*
Left caudate and putamen	52	[-14, 8, 6] *[-10, 0, 8]* *[-20, 2, -4]*

Table S2: Significant clusters of pseudoword > word contrast collapsing across all subjects (primary $p < .001$, Gaussian Random Field cluster-size threshold $p_{FWE} < .05$), including the MNI coordinate of the peak voxel and at most two other local maxima (> 8mm distance).

Figure S1. Sagittal slices of the regions that a significantly greater activation for pseudowords than words (red) or a significantly greater activation for words than pseudowords (blue) , using a primary $p < .001$, cluster-size threshold $p_{FWE} < .05$.