The Neural Basis of Successful Word Reading in Aphasia

Sara B. Pillay¹, William L. Gross¹, William W. Graves², Colin Humphries¹, Diane S. Book¹, and Jeffrey R. Binder¹

Abstract

Understanding the neural basis of recovery from stroke is a major research goal. Many functional neuroimaging studies have identified changes in brain activity in people with aphasia, but it is unclear whether these changes truly support successful performance or merely reflect increased task difficulty. We addressed this problem by examining differences in brain activity associated with correct and incorrect responses on an overt reading task. On the basis of previous proposals that semantic retrieval can assist pronunciation of written words, we hypothesized that recruitment of semantic areas would be greater on successful trials. Participants were 21 patients with left-hemisphere stroke with phonologic retrieval deficits. They read words aloud during an event-related fMRI paradigm. BOLD signals obtained during correct and incorrect trials were contrasted to highlight brain activity specific to successful trials. Successful word reading was associated with higher BOLD signal in the left angular gyrus. In contrast, BOLD signal in bilateral posterior inferior frontal cortex, SMA, and anterior cingulate cortex was greater on incorrect trials. These data show for the first time the brain regions where neural activity is correlated specifically with successful performance in people with aphasia. The angular gyrus is a key node in the semantic network, consistent with the hypothesis that additional recruitment of the semantic system contributes to successful word production when phonologic retrieval is impaired. Higher activity in other brain regions during incorrect trials likely reflects secondary engagement of attention, working memory, and error monitoring processes when phonologic retrieval is unsuccessful.

INTRODUCTION

Recovery of language functions after brain damage is often assumed to involve a degree of reorganization of neural processing, but this recovery mechanism is still not well understood. Despite a large number of functional neuroimaging studies documenting perilesional and contralesional activation in patients with aphasia during language tasks (e.g., Mohr et al., 2016; Hamilton, Chrysikou, & Coslett, 2011; Fridriksson, Baker, & Moser, 2009; Thompson & den Ouden, 2008; Crosson et al., 2007; Saur et al., 2006; Heiss, Kessler, Thiel, Ghaemi, & Karbe, 1999), there is continued debate about the nature of this activation and how it contributes to recovery (Lee, Zreik, & Hamilton, 2017; Gainotti, 2015; Geranmayeh, Brownsett, & Wise, 2014; Fridriksson, Richardson, Fillmore, & Cai, 2012; Turkeltaub et al., 2012; Crosson et al., 2007; Hillis, 2005; Price & Crinion, 2005). An inherent confounding factor in such studies is that the tasks used are likely to be more difficult for patients with brain damage than for healthy controls; therefore, any additional activation in patients might simply reflect additional effort or recruitment of attention and working memory resources (Geranmayeh et al., 2014). Furthermore, it is unclear whether such recruitment necessarily improves performance, and there is even evidence that such activation can be a marker of poorer recovery (Postman-Caucheteux et al., 2010; Martin et al., 2009; Naeser et al., 2004; Rosen et al., 2000).

To address more directly the question of how neural activity enhances performance after stroke, we asked whether activation in patients is directly correlated with improved performance. Recovery of language after stroke is almost never an all-or-none phenomenon; instead, patients typically show partial recovery characterized by variable degrees of success across a set of trials. This trial-to-trial variability makes it possible to examine patterns of neural activity that are uniquely associated with correct and incorrect responses.

Performance-correlated brain activity has been explored only rarely in patients with aphasia. Fridriksson and colleagues (2009) compared fMRI signals during correct and incorrect responses on a picture naming task in 11 patients with aphasia. Although qualitative differences were reported between the correct and incorrect trial activation maps obtained from contrasts with a low-level sensory control, no significant differences emerged when correct and incorrect responses were directly contrasted. This null result suggests that any difference in activation between correct and incorrect responses may be subtle. In addition, the sample of patients with aphasia examined was small and had heterogeneous behavioral profiles; thus, power was likely reduced for finding a significant result (Fridriksson et al., 2009). Postman-Caucheteux

¹Medical College of Wisconsin, ²Rutgers University

© 2017 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 30:4, pp. 514–525
doi:10.1162/jocn_a_01214
et al. (2010) studied three patients with fMRI during picture naming, using a similar event-related analysis to compare activation during correct and incorrect responses. The authors observed relatively greater right frontal activation during incorrect responses but no consistent areas of greater activation during correct trials. Lee et al. (2017) compared correct and incorrect trials during picture naming in a single patient with semantic deficits. In contrast to the right frontal activation reported by Postman-Caucheteux et al., error trials were associated with greater activation in the right superior temporal sulcus and supramarginal gyrus (Lee et al., 2017). Similar to the Postman-Caucheteux et al. results, no areas showed greater activation during correct responses. These results are all in direct contrast to a single case study by Meinzer et al. (2006), who showed that correct responses on a similar picture naming task were associated with greater right frontal activation compared with incorrect responses.

This study uses an overt word reading task to assess performance-correlated brain activity. We reasoned that this task might be more sensitive than picture naming for detecting adaptive brain reorganization because of the availability of alternative processing strategies for phonologic retrieval in the case of words. Whereas picture naming cannot proceed without activation of the concept (semantic representation) that the picture represents, word pronunciation is thought to involve only partial or optional activation of a semantic representation (Graves et al., 2014; Plaut, McClelland, Seidenberg, & Patterson, 1996). Strokes affecting the middle cerebral artery (MCA) territory very often damage perisylvian phonological processing networks while leaving semantic systems relatively intact (Pillay, Stengel, Humphries, Book, & Binder, 2014; Ripamonti et al., 2014; Rapsac et al., 2009). We hypothesized therefore that, in patients with stroke with primarily phonologic system damage, the relatively preserved semantic network might play a role in facilitating word pronunciation. In some well-supported theories of word pronunciation, activation of a word’s meaning (semantic representation) is thought to provide additional information that can assist in retrieving a phonologic code (Plaut et al., 1996). Activation in undamaged regions of the semantic network (Binder, Desai, Graves, & Conant, 2009) was therefore predicted to be greater on trials in which pronunciation was successful compared with trials in which pronunciation was incorrect.

METHODS

Participants

Participants were 21 patients with chronic left-hemisphere ischemic stroke (10 women, 11 men) who demonstrated a residual phonologic retrieval deficit together with intact semantic processing. Participants were selected based on their behavioral deficit from a group of patients with chronic aphasia enrolled prospectively in a larger study. All patients with this behavioral profile were eligible, regardless of where in the left hemisphere the lesion was located. To determine phonologic and semantic function, patients were tested on behavioral tasks (described below) before MRI scanning. Impaired phonologic processing was defined as a z score less than −1.67 (critical value for bottom tail of distribution, $p = .05$) on a pseudoword rhyme matching task, and intact semantic processing was defined as a z score greater than −1.67 on a semantic matching task. Raw scores were transformed to z scores based on performance data from 24 healthy, right-handed, age- and education-matched controls (mean age = 57.7 years, mean education = 15 years).

All patients were at least 180 days poststroke, native English speakers, and premorbidly right-handed according to the Edinburgh Handedness Inventory handedness quotient ($M = 84.0, SD = 25.9$; Oldfield, 1971). Lesions included 17 MCA infarcts, 2 combined MCA/anterior cerebral artery infarcts, and 2 combined MCA/posterior cerebral artery infarcts. Demographic data for patients are listed in Table 1, and individual participant data are presented in Table 2. All patients provided written informed consent to participate and were enrolled prospectively under a protocol approved by the Medical College of Wisconsin institutional review board and undertaken in accord with the Declaration of Helsinki.

Screening Language Assessment

Behavioral tasks were given to characterize key language abilities for inclusion in the study before fMRI scanning. The tasks were administered on a laptop computer connected to a touch-sensitive LCD monitor (ELO Model 1522L; Boston, MA) and programmed in the “LiveCode” environment (https://livecode.com). Testing occurred in a quiet clinic room. Participants initiated each trial by touching an empty green square on the computer screen.

To measure phonologic impairment, participants were asked to complete a 72-trial pseudoword rhyme matching task (Figure 1, left). On each trial, a sample

| Table 1. Patient Demographic and Performance Data |
|---------|--------|-------|----------|
| Variable               | Mean   | SD    | Range    |
| Age (years)            | 56.4   | 12.5  | 30–80    |
| Education (years)      | 14.7   | 3.2   | 8–20     |
| Days post onset        | 1,134  | 1,491 | 180–6,732 |
| Lesion size (mL)       | 73,439 | 58,567| 6,711–226,978 |
| Semantic picture matching | 93.8  | 1.7   | 90.8–96.7 |
| Pseudoword rhyme matching | 67.5  | 14.1  | 45.8–87.5 |
pseudoword (e.g., boark) was presented in the center of a computer display with two choice items (e.g., lorge and rauke) presented side by side below the sample. Participants indicated via button press which of the two items rhymed with the sample. Trials were designed to take advantage of the fact that English has multiple possible spellings for the same sound (e.g., vair » plare/sbar), enabling decoupling of phonologic similarity from orthographic similarity. All items had unambiguous (i.e., highly consistent) pronunciations.

To assess semantic processing, participants completed a 120-trial semantic picture matching task, similar to the Camel and Cactus test (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000; Figure 1, right). This task required participants to select from two choices the item that was most similar to the sample picture (i.e., eagle » robin vs. back).

Patients also completed an 80-item picture naming task described elsewhere (Pillay, Binder, Humphries, Gross, & Book, 2017). Results are provided in Table 2 to further characterize the aphasia severity in each patient but were not used in the context of this study and are not discussed further.

### Table 2. Individual Patient Demographic and Performance Data

<table>
<thead>
<tr>
<th>Pt.</th>
<th>Age (years)</th>
<th>Educ. (years)</th>
<th>DPO</th>
<th>PI</th>
<th>No. of Lesioned Voxels</th>
<th>Picture Naming</th>
<th>Semantic Picture Matching</th>
<th>Pseudoword Rhyme Matching</th>
<th>Word Reading (fMRI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947</td>
<td>55</td>
<td>12</td>
<td>3,053</td>
<td>−6.57</td>
<td>87,201</td>
<td>65.0</td>
<td>94.2</td>
<td>62.5</td>
<td>76.4</td>
</tr>
<tr>
<td>2471</td>
<td>80</td>
<td>17</td>
<td>658</td>
<td>−9.81</td>
<td>46,937</td>
<td>73.8</td>
<td>93.3</td>
<td>45.8</td>
<td>79.2</td>
</tr>
<tr>
<td>2477</td>
<td>30</td>
<td>14</td>
<td>961</td>
<td>−2.84</td>
<td>85,388</td>
<td>95.0</td>
<td>96.7</td>
<td>81.7</td>
<td>90.3</td>
</tr>
<tr>
<td>3315</td>
<td>49</td>
<td>14</td>
<td>255</td>
<td>−1.72</td>
<td>24,230</td>
<td>95.0</td>
<td>92.5</td>
<td>87.5</td>
<td>86.1</td>
</tr>
<tr>
<td>3326</td>
<td>52</td>
<td>20</td>
<td>210</td>
<td>−1.72</td>
<td>10,416</td>
<td>93.8</td>
<td>94.2</td>
<td>87.5</td>
<td>94.4</td>
</tr>
<tr>
<td>3380</td>
<td>61</td>
<td>19</td>
<td>726</td>
<td>−2.26</td>
<td>32,580</td>
<td>100.0</td>
<td>95.0</td>
<td>84.7</td>
<td>97.2</td>
</tr>
<tr>
<td>3383</td>
<td>54</td>
<td>14</td>
<td>1,166</td>
<td>−8.46</td>
<td>130,491</td>
<td>66.3</td>
<td>91.7</td>
<td>52.8</td>
<td>63.9</td>
</tr>
<tr>
<td>3390</td>
<td>44</td>
<td>12</td>
<td>253</td>
<td>−1.99</td>
<td>72,010</td>
<td>93.8</td>
<td>95.0</td>
<td>86.1</td>
<td>97.2</td>
</tr>
<tr>
<td>3584</td>
<td>52</td>
<td>12</td>
<td>350</td>
<td>−4.14</td>
<td>74,148</td>
<td>46.3</td>
<td>95.0</td>
<td>75.0</td>
<td>51.4</td>
</tr>
<tr>
<td>3588</td>
<td>71</td>
<td>14</td>
<td>251</td>
<td>−6.57</td>
<td>23,355</td>
<td>27.5</td>
<td>92.5</td>
<td>62.5</td>
<td>27.8</td>
</tr>
<tr>
<td>3590</td>
<td>62</td>
<td>16</td>
<td>934</td>
<td>−5.49</td>
<td>33,329</td>
<td>58.8</td>
<td>91.7</td>
<td>68.1</td>
<td>81.9</td>
</tr>
<tr>
<td>3603</td>
<td>62</td>
<td>14</td>
<td>449</td>
<td>−6.84</td>
<td>71,445</td>
<td>80.0</td>
<td>91.7</td>
<td>61.1</td>
<td>69.4</td>
</tr>
<tr>
<td>3666</td>
<td>77</td>
<td>16</td>
<td>638</td>
<td>−9.54</td>
<td>18,597</td>
<td>77.5</td>
<td>94.2</td>
<td>47.2</td>
<td>80.6</td>
</tr>
<tr>
<td>3820</td>
<td>40</td>
<td>15</td>
<td>6,732</td>
<td>−6.57</td>
<td>156,674</td>
<td>90.0</td>
<td>95.0</td>
<td>62.5</td>
<td>93.1</td>
</tr>
<tr>
<td>4091</td>
<td>57</td>
<td>12</td>
<td>2,495</td>
<td>−1.72</td>
<td>68,337</td>
<td>98.8</td>
<td>95.8</td>
<td>87.5</td>
<td>95.8</td>
</tr>
<tr>
<td>4229</td>
<td>64</td>
<td>8</td>
<td>233</td>
<td>−6.03</td>
<td>95,698</td>
<td>78.8</td>
<td>90.8</td>
<td>65.3</td>
<td>68.1</td>
</tr>
<tr>
<td>4572</td>
<td>38</td>
<td>18</td>
<td>1,278</td>
<td>−5.49</td>
<td>226,978</td>
<td>31.3</td>
<td>92.5</td>
<td>68.1</td>
<td>38.9</td>
</tr>
<tr>
<td>6887</td>
<td>55</td>
<td>14</td>
<td>1,175</td>
<td>−7.91</td>
<td>184,005</td>
<td>28.8</td>
<td>95.0</td>
<td>55.6</td>
<td>45.8</td>
</tr>
<tr>
<td>8342</td>
<td>47</td>
<td>15</td>
<td>180</td>
<td>−9.27</td>
<td>45,489</td>
<td>83.8</td>
<td>92.5</td>
<td>48.6</td>
<td>70.8</td>
</tr>
<tr>
<td>8754</td>
<td>68</td>
<td>13</td>
<td>214</td>
<td>−7.91</td>
<td>48,220</td>
<td>61.3</td>
<td>92.5</td>
<td>55.6</td>
<td>91.7</td>
</tr>
<tr>
<td>9720</td>
<td>57</td>
<td>14</td>
<td>605</td>
<td>−4.69</td>
<td>6,711</td>
<td>100.0</td>
<td>96.7</td>
<td>72.2</td>
<td>93.1</td>
</tr>
</tbody>
</table>

DPO = days post onset; PI = phonological impairment (z scores).

Figure 1. Example trials from the behavioral tasks given for participant inclusion outside the MRI environment. The left image shows an example of the pseudoword rhyme matching task. The right image shows an example of the semantic picture matching task.
Functional Imaging

fMRI Tasks

Participants who met inclusion criteria on the language screening assessment then underwent an fMRI session, during which they performed an overt reading task with single words. Spoken responses were used because they allowed for definitive scoring of responses and classification of correct and incorrect trials. Participants read aloud 72 concrete nouns ranging in length from four to six characters. Imageability (ease of mental image retrieval, rated on a 1–7 scale; Cortese & Fugett, 2004) ranged from 3.66 to 6.43, with a mean of 5.50. Written word frequency ranged from 8 to 260 per million, with a mean of 59 per million (Baayen, Piepenbrock, & Gulikers, 1995). In addition to the reading trials, 72 fixation intervals were randomly dispersed among the trials to create a variable ISI for optimal fast event-related analysis.

Stimulus Presentation and Response Recording

Timing and order of stimulus presentation were controlled with E-Prime software (Psychology Tools Inc., Pittsburgh, PA) and synchronized to MRI data acquisition. Standard 800 × 600 pixel RGB output was sent from a PC to an LCD video projector. Images were back-projected on a screen located 240 cm from a prism lens mounted within the head coil just above the participant’s eyes. Individual stimuli were presented in the center of a screen in white font on a black background and subtended a visual angle of 2.0°–2.5°. Stimuli were presented for 2000 msec and followed by a fixation cross until the next stimulus. The mean trial duration was 6.58 sec.

Spoken responses were captured via a fiber-optic dual-channel noise-canceling microphone (FOMRI-III; Opto-acoustics, Mazor, Israel) positioned near the participant’s mouth. Tones marking the onset of each trial (inaudible to the participant) were recorded on a second audio channel. Responses were then phonetically transcribed offline for accuracy, RT, and error analyses. Responses were considered errors if the participant failed to produce a complete response or failed to produce a correct response. Incomplete or incorrect responses that were followed immediately by correct responses were counted as correct as long as the correct response occurred before the next trial. For completeness, error responses were also subclassified as omissions (i.e., no response or fragmentary utterance), phonologic paraphasias (i.e., phonologically related neologism or word), semantic paraphasias, or phonologically and semantically unrelated errors, although these error types were not separately analyzed in the fMRI analysis.

Word reading is often a challenging task for people with aphasia and becomes more challenging when coupled with performing in an unfamiliar fMRI environment with constraints on available RT. Functional neuroimaging measurements are sensitive to RT, working memory, and attentional effort (Binder, Medler, Desai, Conant, & Liebenthal, 2005), and the use of an overt production task permits collection of RTs for assessment of overall trial difficulty. A custom MATLAB (The MathWorks, Natick, MA) script was used to estimate RT, from stimulus onset to response onset, for each trial in each participant. Each auditory waveform and response onset marker was visually inspected for accuracy and manually adjusted as needed. These RT data were incorporated directly into the image analysis in an attempt to delineate brain areas modulated by task difficulty (Binder, Medler, et al., 2005). Two patients did not have RT data available because of error in collecting stimulus onset times, reducing the n to 19 for this additional analysis.

MRI Acquisition

All fMRI data were acquired at 3 T using a GE (Chicago, IL) Excite whole-body scanner. Four patients were scanned with a 32-channel head coil. The remaining participants were scanned with an eight-channel head coil. A gradient-echo, EPI sequence (flip angle = 77°, echo time = 25 msec, repetition time = 2000 msec, number of excitations = 1, field of view = 192 cm, matrix = 64 × 64, slice thickness = 3 mm, interslice gap = 0.5 mm, 36 axial slices) was used for isotropic (3 × 3 × 3 mm) whole-brain fMRI data acquisition. Head movement was monitored during scanning using realtime image registration. High-resolution T1-weighted anatomical reference images were obtained using a spoiled gradient-recalled sequence (flip angle = 12°, echo time = MinFull, T1 [prep] = 450 msec, repetition time = 8.2 msec, NEX = 1, voxels = 1.0 × 1.0 × 1.0 mm, 162 axial slices).

Lesion Tracing and Template Registration

Lesioned areas were identified using a semiautomated procedure to create an individual lesion map for each patient (Pillay et al., 2014). Each patient’s anatomical image and associated lesion map were then morphed to a stereotaxic template (Colin N27) using 3dQwarp, a nonlinear registration routine in the Analysis of Functional NeuroImages (AFNI) software package (https://afni.nimh.nih.gov). The registration included constrained cost-function masking using the lesion volume as a mask and resampling to a nominal 1 × 1 × 1 mm voxel grid. The parameters used for warping each individual’s anatomy to the template were then applied to the functional images. This nonlinear registration process corrects for anatomical distortions that are common after focal brain damage, particularly local ventricular enlargement. The group lesion overlap map is presented in Figure 2.

fMRI Data Analysis

Image processing and statistical analyses were performed using AFNI software. Preprocessing steps included slice timing correction and image registration.
For each participant, the first four images in each time series were discarded before regression analysis to avoid saturation effects. Translation and rotation parameters, estimated during registration, were saved for use as noise covariates. EPI volumes were then registered to the T1 anatomical image using a modality-specific cost function based on weighted local Pearson coefficients (Saad et al., 2009) using the AFNI script \texttt{align_epi_anat.py}. Images contaminated by large artifacts, such as motion, were detected automatically using regression techniques built into the AFNI program \texttt{3dToutcount} and censored from the analysis. No data sets required censoring of more than 10% of the images (mean percent censored = 3.78, \(SD = 2.11\)).

BOLD signal changes were analyzed with a multiple linear regression model using the AFNI program \texttt{3dDeconvolve}. Stimulus regressors were created for each task condition, convolved with a canonical gamma variate hemodynamic response function. Six motion vectors computed during image registration were included as covariates of no interest, along with an additional covariate from signal in the ventricles to estimate speech movement artifact (Graves, Grabowski, Mehta, & Gordon, 2007). A Gaussian kernel of 5 mm FWHM was used for smoothing before the regression analysis.

**Group Contrasts**

The contrast of interest was between words that were read successfully and words that elicited errors (Correct − Error), which specifically identifies brain activity correlated with successful performance. Because of variable patient performance and therefore unequal numbers of trials in each condition across patients, it was necessary to determine the lower limit of trials acceptable for participant inclusion in the analysis. As the number of trials for a particular condition (correct or incorrect) decreases, the reliability of the corresponding parameter estimate also decreases (i.e., the standard deviation of the estimate increases). Increasing the criterion for minimum trial number, however, reduces power at the group level because of the reduction in number of participants who meet criteria. Model quality for a given combination of minimum trial number and group sample size can be determined by inverting the regressor matrix and taking the standard deviation of the resulting values. In formal terms, the linear regression model for this analysis is given by

\[ y = X\beta + \epsilon, \]

where \(y\) is the measured response, \(X\) is the design matrix, \(\beta\) is the vector of unknown parameters, and \(\epsilon\) is the measurement error. The least squares estimate of \(\beta\) is given by

\[ b = (X^T X)^{-1}X^T y, \]

and the variance of the parameter estimate is estimated by

\[ s^2(b) = \text{MSE}(X^T X)^{-1}, \]

where MSE is the mean square error, a scalar that estimates the error measurement variance. The standard deviation for a particular element of the \(b\) vector, \(s(b_k)\), is given by the square root of the corresponding diagonal element of the \(s^2(b)\) matrix. If the measurement variance is assumed to be constant across experimental designs, then \(s^2(b)\) is a function of the structure of \(X\) alone, allowing computation of a normalized standard deviation (nSD) for each experimental design (i.e., each combination of minimum trial number criterion and resulting group sample size). Minimizing this nSD produces the most reliable final parameter estimates (Ward, 2006). Using this method, nSD was calculated by setting a minimum trial number criterion and using the observed distribution of trial numbers in the real data to determine the resulting group sample size. This procedure was repeated over a range of minimum trial number criteria. A minimum of two time points for participant inclusion was determined to be optimal for minimizing nSD and thus maximizing the reliability of the parameter estimates and power at the group level. Thus, patients were included in the analysis if they made at least two error responses.

The result of this simulation may seem surprising, given that a sample of two trials would be very unlikely to yield a reliable activation map at the individual participant level. The crucial point here is that there is a direct trade-off between minimum number of trials allowed and sample size at the group level. The simulation indicates that sample size at the group level has a much larger effect than power at the individual participant level on power at the group level. Most of the sample (62%) made more than 10 error responses. Two patients made only two error responses, yet the simulation suggests that power at the group level is enhanced when these patients are included. To test this prediction, we repeated the main contrast between correct and incorrect trials but included only patients with at least 10 errors, which reduced the sample size to 13 patients. The results were similar; however, they did not survive significance, likely because of reduced power.

Patients were selected for inclusion if they had a phonologic access deficit and relatively preserved semantic store. We hypothesized that the preserved semantic
network may play a role in facilitating word production (Plaut et al., 1996) and that activation in undamaged regions of the semantic network may be greater for correct trials than incorrect trials. The overlap between the thresholded correct–error activation map and semantic regions identified in a previous meta-analysis (Binder et al., 2009) was examined to determine whether patients engaged semantic regions more for correct than incorrect trials.

Two additional analyses were completed in an effort to reduce variance due to individual factors. In the first analysis, the phonologic impairment z score for each patient was included as a covariate to remove variation correlated with impairment severity. In theory, the severity of phonologic impairment might systematically modulate the degree of compensatory activation in undamaged brain regions, thus adding between-participant variance. Controlling for this factor might enhance detection of activation that is independent of severity level. In the second analysis, lesioned tissue was excluded on a voxel-wise basis from the group level t test. That is, the group level t test at each voxel included only the subset of patients who did not have a lesion involving the voxel. Presumably, this reduces variance at each voxel resulting from the presence or absence of a lesion, which may improve sensitivity despite the smaller sample size at some voxels. This method results in the degrees of freedom varying across voxels, depending on the group level lesion burden (e.g., a voxel in the center of the left insula would have fewer degrees of freedom because it is more likely to be lesioned than a voxel in the left occipital lobe). Each voxel was thresholded, however, to the same alpha level (p < .01).

Finally, the RTs on each trial (including error trials) were included in a regression model, after normalization of the RT values, to account for variance due solely to time-on-task, a presumed marker of task difficulty. Separate RT maps were created for correct and incorrect conditions, and these maps were then combined within participants so that the resulting RT map captured variance independent of condition. As there were different numbers of trials in each condition, within-participant RT maps were weighted by the number of trials per condition before being averaged.

All contrasts were first computed at the individual participant level. Beta coefficient maps for these contrast maps (registered to the common template space) were then analyzed with a single-sample t test at the group level. Activation maps were thresholded at voxel-wise p < .01 and a minimum cluster size of 1609 µL, corresponding to a whole-brain p < .05. The whole-brain probability was derived from Monte Carlo simulations using simulated data. Spatial smoothing was applied to the simulated data using a model composed of a Gaussian and monoexponential function, which was fit to the residual data set from the regression analysis (using 3dFWHMx in AFNI). The addition of the monoexponential function to the Gaussian model was performed to address the increased false-positive rates found when using a Gaussian model alone (Cox, Chen, Glen, Reynolds, & Taylor, 2017; Eklund, Nichols, & Knutsson, 2016).

RESULTS

Task Performance

Performance on the two behavioral measures completed outside the MRI environment is reported in Table 1. Across all patients, semantic matching performance (M = 93.8, SD = 1.7) was significantly better than pseudoword rhyme matching performance (M = 67.5, SD = 14.1), as prescribed by the inclusion criteria. Patients performed worse than controls on the pseudoword rhyme matching task, t(37) = −7.89, p < .001, but did not perform differently from controls on the semantic matching task, t(37) = −1.43, p = .16.

Mean accuracy on the fMRI word reading task was 75.9% (SD = 20.4%). The number of errors ranged from 2 (2.8%) to 52 (72.2%). Words read correctly were significantly more imageable (M = 5.56, SD = 0.06) than words read incorrectly (M = 5.14, SD = 0.28; p < .001). In addition, words read correctly had higher frequencies (M = 64.2, SD = 3.3) than words read incorrectly (M = 41.2, SD = 13.1; p < .001). Most errors were phonologically related to the target item (M = 10.4, SD = 8.2), followed by errors that were unrelated to the target item (M = 3.7, SD = 6.7). Semantically related errors were rare (M = 1.7, SD = 1.4), as were omissions (M = 1.4, SD = 2.2). Reading accuracy was not correlated with age, education, gender, or days post onset. Individual patient performances are shown in Table 2.

Mean RT on the reading task was 1291 msec (SD = 380 msec) across all trials. Correct responses (M = 1228 msec, SD = 356 msec) were faster than incorrect responses (M = 1362 msec, SD = 399 msec), but this difference was not statistically significant (p = .469). Across participants, mean RT for correct trials was negatively correlated with reading accuracy (r = −.50, p = .03).

fMRI Results

Stereotaxic coordinates of activation peaks are listed in Table 3. Areas that were differentially engaged during successful and unsuccessful reading are shown in Figure 3A. The left angular gyrus (AG) was more active for correct word trials (correct > error). In contrast, four regions were more active for incorrect word trials (error > correct). The largest of these were located bilaterally in the SMA and dorsal anterior cingulate gyrus. The other two clusters involved similar regions in the right and left posterior inferior frontal cortex, extending from the inferior precentral gyrus to the pars opercularis of the inferior frontal gyrus (IFG). On the right side, this cluster extended farther ventrally into the anterior insula.
Brain areas modulated by RT regardless of stimulus condition are shown in Figure 3B. Several regions identified in the error > correct contrast also showed positive correlations with time-on-task, including the SMA and dorsal anterior cingulate cortex bilaterally, the right posterior IFG and precentral gyrus, and the right anterior insula. Other areas modulated by RT included the right central sulcus, bilateral mid-cingulate and posterior cingulate gyrus, bilateral cuneus and lingual gyrus, and bilateral thalamus. No areas were negatively correlated with RT.

To assess the relationship between the left AG area activated during successful reading and areas previously identified with semantic processing, the correct–error activation map was overlapped with a meta-analysis map of regions activated by semantic contrasts in healthy participants (Binder et al., 2009). As shown in Figure 4, the left AG region associated with correct reading performance in patients overlapped completely with the previously identified semantic network, whereas only a minute portion of the error > correct map overlapped with semantic areas in the left IFG.

Inclusion of phonological impairment z score as a covariate at the group level had no notable effect on the results. Similarly, exclusion of lesioned voxels, which constrains the analysis at each voxel to only those patients

**Table 3. Peak Coordinates within Significantly Activated Clusters**

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Location of Extreme Point</th>
<th>Cluster size (μL)</th>
<th>Talairach Coordinates</th>
<th>z Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>L AG</td>
<td>2,430</td>
<td>−40</td>
<td>−52</td>
</tr>
<tr>
<td>Incorrect</td>
<td>R IFG</td>
<td>5,388</td>
<td>38</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>L SMA</td>
<td>5,833</td>
<td>−6</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>L IFG</td>
<td>3,471</td>
<td>−36</td>
<td>0</td>
</tr>
</tbody>
</table>

L = left; R = right.

**Figure 3.** Group z score contrasts. (A) Correct–error. The area more active for correctly read words (warm colors) was in the left AG. Areas more active for incorrect words (cool colors) were located bilaterally in the posterior inferior frontal cortex (mainly precentral gyrus and pars opercularis) and posterior medial frontal cortex (mainly SMA). (B) Areas modulated by RT including the bilateral SMA, posterior precuneus and lingual gyrus, and right IFG and adjacent precentral gyrus/sulcus.
DISCUSSION

This study shows for the first time the brain regions where neural activity is correlated specifically with successful word reading in people with aphasia. By contrasting the brain states present during correct and incorrect responses, the analysis distinguishes activation supporting successful phonologic retrieval from nonspecific activation related to increased effort. Activation associated with correct word reading responses was located in the left AG. In contrast, activation associated with incorrect responses was located mainly in regions linked with attention and working memory processes (Hampshire, Chamberlain, Monti, Duncan, & Owen, 2010; Binder, Medler, et al., 2005; Owen, McMillan, Laird, & Bullmore, 2005), including bilateral SMA and dorsal anterior cingulate gyrus, posterior IFG and precentral gyrus, and anterior insula.

The left AG has been strongly implicated in semantic processing in healthy participants and was the most consistently activated region in a meta-analysis of 120 functional imaging studies using semantic task contrasts (Binder et al., 2009). The left AG has also been linked with semantic processing during reading tasks in healthy controls, showing positive modulatory effects of imageability (Graves, Desai, Humphries, Seidenberg, & Binder, 2010; Binder, Medler, et al., 2005; Binder, Westbury, McKiernan, Possing, & Medler, 2005) and word frequency (Graves et al., 2010; Prabhakaran, Blumstein, Myers, Hutchison, & Britton, 2006). Word imageability (the ease with which a word evokes an associated mental image) has a weak facilitatory effect on RT in overt reading tasks (Woollams, 2005; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Strain, Patterson, & Seidenberg, 1995), suggesting that retrieving word meaning can speed computation of phonology. This idea is made explicit in the “triangle model” of word reading, in which phonologic units receive input from both orthographic and semantic representations (Plaut et al., 1996). Graves et al. (2014) showed that the degree to which healthy participants access semantic information in reading aloud, as indexed by the size of their imageability facilitation effect, is correlated with the volume of the white matter pathway connecting the left AG with the posterior left superior temporal gyrus, a region implicated in phonologic retrieval (Pillay et al., 2014; Graves, Grabowski, Mehta, & Gupta, 2008). Together, these computational and empirical studies suggest that activation of semantic information can facilitate phonologic retrieval in healthy readers and that one neurobiological mechanism for this facilitation is input from the left AG to the perisylvian phonologic system. Other ancillary support for this model comes from a performance-based study in healthy participants examining neural activity for naming famous faces, in which activation in the left AG (Talairach coordinates = −42, −56, 47; Euclidean distance from peak AG point in this study = 6 mm) was associated with greater activation for correct compared with incorrect naming (Gesierich et al., 2012), suggesting that increased semantic access for proper nouns also boosts phonologic retrieval.

The present results are consistent with this “semantic boost” mechanism in two ways. First, the left AG was more active when correct responses were produced, suggesting that phonologic retrieval was facilitated by AG activation. Second, correctly pronounced words were, on average, higher in imageability and frequency than incorrectly pronounced words. These factors index the ease with which word meaning is retrieved (Graves et al., 2010; Binder, Westbury, et al., 2005; Paivio, 1990; Monsell, Doyle, & Haggard, 1989). Thus, the results suggest that phonologic retrieval in chronic aphasia is more likely when word meaning is successfully activated.

Although we have interpreted the left AG activation as reflecting more extensive semantic retrieval during correct trials, an alternative possibility is that this region was more extensively deactivated during incorrect trials. The concept of deactivation refers to a relative decrease in neural activity compared with a putative baseline “resting” state; however, it is now widely accepted that the condition called “resting” is actually a cognitively active state involving “mind wandering,” reactivation of episodic memories, future planning, and other mental activities that require complex semantic processing (Andrews-Hanna, Saxe, & Yarkoni, 2014; Spreng & Grady, 2010; Binder et al., 1999, 2009; Buckner, Andrews-Hanna, & Schacter, 2008; Mason et al., 2007; McKiernan,
IFG activation during picture naming could reflect a degree of deactivation. A second weakness of the alternative model is that it offers no alternative account of why some deactivations in fMRI studies represent, at least in part, an interruption of these ongoing conceptual processes, and it is therefore possible that purported differences in activation between stimulus conditions in the left AG instead reflect differences in deactivation, without the need to postulate differences in semantic processing of the stimuli. Given the low temporal resolution of BOLD fMRI measurements, this method cannot distinguish between deactivation followed by different degrees of activation versus different degrees of initial deactivation. We believe that the latter model is unlikely for several reasons. First, the postulated mechanism underlying variation in deactivation is variation in time-on-task. That is, a longer processing time would be expected to cause longer interruption of ongoing conceptual processes and a larger decrease in the BOLD signal after integration of the signal over time. In contrast to this prediction, we observed no areas where BOLD signal was negatively correlated with RT. This result is consistent with other fMRI studies of single-word reading that found little or no negative correlation between RT and BOLD signal, even for RT variations spanning hundreds of milliseconds (Graves et al., 2010; Binder, Medler, et al., 2005; Binder, Westbury, et al., 2005). This evidence suggests that, at least on the time scale of a single-word reading task, variation in difficulty is not associated with variation in the degree of deactivation. A second weakness of the alternative model is that it offers no alternative account of why some trials lead to successful phonological retrieval. If anything, greater suppression of “default mode” activity has been reported to improve stimulus processing rather than the converse (Estes, Noonan, Rosenberg, & Degutis, 2013; Smallwood & Schooler, 2006; Weissman, Roberts, Visscher, & Woldorff, 2006). Finally, the alternative model offers no account of how differences in imageability and word frequency between successful and unsuccessful trials affect brain activation or phonologic retrieval.

To our knowledge, the only previous report of enhanced brain activation during correct compared with incorrect responses in chronic aphasia was the single case studied by Meinzer et al. (2006). Correct picture naming in that case was associated with greater activation in the right IFG, with the strongest effects in the anterior IFG (pars orbitals). The results thus differ from this study in linking correct response activation to an entirely different region (right IFG vs. left AG) and in the direction of the effect observed in right IFG (stronger for correct responses in Meinzer et al. vs. stronger for error responses in this study). Several factors could account for the conflicting findings. Unlike reading aloud, picture naming necessarily requires selection of a lexical label from competing alternatives before phonologic retrieval. The left anterior IFG has been linked with lexical selection in a number of fMRI studies (Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997). Thus, right anterior IFG activation during picture naming could reflect a transfer of this selection mechanism to the right IFG in this patient. Another factor is that the current study is a composite or general result derived from a large sample of patients; thus, the results capture effects that are common across individuals in the sample. It is possible that the right IFG effect observed by Meinzer et al. reflected the particular lesion pattern or premorbid language organization of their patient, and it is unknown whether this effect would be reliable across a sample. Finally, the region of right IFG implicated in the two studies is different, in that the positive effects observed by Meinzer et al. were in the anterior aspect of the gyrus, whereas the negative effects we observed were in the posterior IFG and precentral gyrus. Thus, different segments of the right IFG may have different functions in chronic aphasia recovery.

The results of our error > correct comparison are in general agreement with another previous performance-based fMRI study by Postman-Caucheteux and colleagues (2010), which used an overt picture naming task. Similar to our results, these authors observed activation associated with error responses in a region of the posterior right IFG in three patients with aphasia from left MCA stroke. Together, these results suggest that posterior right IFG activation in chronic aphasia may be a marker of effort rather than a neural correlate of successful phonologic reorganization. Several lines of evidence provide further support for this suggestion. First, the right IFG has been implicated in working memory and cognitive control processes in many prior functional imaging studies of healthy participants (Aron, Robbins, & Poldrack, 2014; White et al., 2014; Hampshire et al., 2010; Owen et al., 2005). Second, the same posterior right IFG region associated with error responses in this study also showed a positive correlation with time-on-task, suggesting sensitivity to item difficulty. This region has also shown positive correlations with RT in several prior reading studies in healthy participants (Taylor, Rastle, & Davis, 2014; Graves et al., 2010; Binder, Medler, et al., 2005; Binder, Westbury, et al., 2005). Third, suppression of the right IFG with TMS has been shown to produce improved picture naming ability in people with chronic aphasia (Naesser et al., 2011), suggesting that activation in the right inferior frontal cortex represents maladaptive effort or maladaptive reorganization of search and selection processes when left perilesional areas are insufficient (Hillis, 2005).

In addition to the right posterior IFG, several other areas showed stronger activation during error responses, and most of these were also positively correlated with time-on-task. All of these regions, including the left posterior IFG, right anterior insula, bilateral precentral gyrus, bilateral SMA, and bilateral anterior cingulate gyrus, have been implicated in attention and cognitive control processes (i.e., working memory, selection, decision, error monitoring; Owen et al., 2005), and all have shown positive correlations with RT in previous reading studies in healthy controls (Taylor et al., 2014; Graves et al., 2010; Binder, Medler, et al., 2005; Binder, Westbury, et al.,
phonology and preserved semantic access. Interactive
tients with a similar processing deficit, that is, impaired
made regularization errors in reading irregular words
anterior parietal phonologic network when patients
et al. (2009) of five patients with semantic dementia. The
authors observed a trend toward greater activation of the
entirely different patterns of compensation in patients
of the phonologic system has simply “recruited” cortex rather
than to semantic mediation. Although possible, this
model provides no account of why words read correctly
in this study tend to be more imageable, a finding that
clearly links the underlying activation with retrieval of
word meaning.

Acknowledgments
We thank the participants for their contribution and time, David
Medler for assistance with experimental design, and B. Douglas
Ward for consultation regarding fMRI analysis. This study was
supported by grants from the NIH National Institute of Neuro-
logical Disorders and Stroke (R01 NS053576, R01 DC003681, and
R03 NS054958) to J. R. B. and by an award from the American
Heart Association (13PRE16510003) to S. B. P. S. B. P. was sup-
ported by a grant from the AHA. J. R. B. and W. W. G. were sup-
ported by grants from the NIH.

Reprint requests should be sent to Sara B. Pillay, Depart-
ment of Neurology, Medical College of Wisconsin, 8701
Watertown Plank Rd., Milwaukee, WI 53226, or via e-mail:
sapillay@mcw.edu.

REFERENCES
of episodic retrieval and mentalizing to autobiographical
thought: Evidence from functional neuroimaging, resting-state
and the right inferior frontal cortex: One decade on. Trends
in Cognitive Sciences, 18, 177–185.
Wagner, A. D. (2005). Dissociable controlled retrieval and
generalized selection mechanisms in ventrolateral prefrontal
Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler,
Where is the semantic system? A critical review and meta-
analysis of 120 functional neuroimaging studies. Cerebral
Cortex, 19, 2767–2796.
Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F.,
during the conscious resting state: A functional MRI study.
Journal of Cognitive Neuroscience, 11, 80–93.
Binder, J. R., Medler, D. A., Desai, R., Conant, L. L., &
Lieberth, E. (2005). Some neurophysiological constraints on
models of word naming. NeuroImage, 27, 677–693.
Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., &
concrete and abstract concepts. Journal of Cognitive
Neuroscience, 17, 905–917.
anoma in two fluent aphasia syndromes. American Journal of
Speech-Language Pathology, 13, 236–249.
Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., &
Hodges, J. R. (2000). Non-verbal semantic impairment in
semantic dementia. Neuropsychologia, 38, 1207–1215.
(2008). The brain’s default network: Anatomy, function, and


