Simulation of Aphasic Naming Performance in Non-Brain-Damaged Adults

JoAnn P. Silkes*
Malcolm R. McNeil**
University of Wisconsin–Madison
Mathias Drton
University of Washington, Seattle

Discussion abounds in the literature as to whether aphasia is a deficit of linguistic competence or linguistic performance and, if it is a performance deficit, what are its precise mechanisms. Considerable evidence suggests that alteration of nonlinguistic factors can affect language performance in aphasia, a finding that raises questions about the modularity of language and the purity of linguistic mechanisms underlying the putative language deficits in persons with aphasia. This study investigated whether temporal stress plus additional cognitive demands placed on non-brain-damaged adults would produce aphasic-like performance on a picture naming task. Two groups of non-brain-damaged participants completed a picture naming task with additional cognitive demands (use of low frequency words and making semantic judgments about the stimuli). A control group performed this task at their own pace, and an experimental group was placed under time constraints. Naming errors were identified and coded by error type. Errors made by individuals with aphasia from a previous study (S. E. Kohn & H. Goodglass, 1985) were recoded with the coding system used in the present study and were then compared with the types of errors produced by the 2 non-brain-damaged groups. Results generally support the hypothesis that the language performance deficits seen in persons with aphasia exist on a continuum with the language performance of non-brain-damaged individuals. Some error type differences between groups warrant further investigation.

KEY WORDS: aphasia, language functions and disorders, anomia, verbal expression

All explanations for the linguistic deficits of aphasia following brain damage assume that the underlying neurological insult causes the observed symptoms. These explanations are traditionally attributed to one of two general mechanisms (McNeil, 1982). The first, which assumes some form of linguistic loss (i.e., a deficit based in reduction of linguistic competence), is derived from the belief that the destruction of brain tissue affects a language center where specific linguistic information is stored and is now lost or altered because of the lesion or that a pathway connecting two or more language centers has been disconnected (McNeil, Odell, & Tseng, 1991). This “centers and pathways” explanation for aphasia (Geschwind, 1965) makes it possible to attribute the symptoms to a loss of linguistically necessary computations or to a loss of linguistic representations.

*Currently at the University of Washington.
**Currently at the University of Pittsburgh and at the VA Pittsburgh Health Care System.
An alternative explanation proposes some form of linguistic inefficiency or deficit in accessing or activating language rules and representations (i.e., a deficit based in reduction of linguistic performance). This performance deficit explanation for aphasia also assumes that the lesion is the ultimate source of the observed language symptoms, but it does not assume a deletion of linguistic information from the aphasic person’s repertoire (McNeil, 1988; McNeil et al., 1991). Rather, it argues that the structure of the linguistic rules and representations remains computationally preserved but is at times less accessible or achievable because of damage to the neural mechanisms that control the activation, selection, and inhibition of linguistic elements. It has been speculated that this difficulty in the accessibility of linguistic procedures could be expected to be shared by non-brain-damaged (NBD) participants under special cognitive, linguistic, affective, or communicative states, especially if these states are similar to the information-processing conditions and demands imposed by the aphasia-producing lesions (McNeil, 1982).

The access deficit model of aphasia (McNeil, 1983, 1988; McNeil et al., 1991) has received considerable support across a variety of language levels and processes, including auditory processing and comprehension (Brookshire & Nicholas, 1984; Miyake, Carpenter, & Just, 1994; Saffran, Schwartz, & Linebarger, 1998; Tseng, McNeil, & Milenkovic, 1993; Wegner, Brookshire, & Nicholas, 1984) and word association (Adams, Reich, & Flowers, 1989). This evidence is further supported by several sources who have shown that, on a variety of tasks, (a) aphasic performance improves with increased processing time and reduced speed of stimulus presentation (Pashek & Brookshire, 1982; Santo Pietro & Rigodsky, 1982), (b) people with aphasia can demonstrate improvements in language performance merely by changing stimulus parameters (e.g., adding tactile information, or making pictures more salient; see Faber & Aten, 1979; McNeil et al., 1991), (c) symptoms identical to aphasia in all ways may be transient in certain conditions (Eisenson, 1984; McNeil, 1988), and (d) persons with aphasia may demonstrate highly variable linguistic performance from moment to moment (McNeil, 1983; Tseng et al., 1993).

It is logical, therefore, that lexical retrieval problems in aphasia should also reflect an access deficit. If this is the case, then it should be possible to induce NBD adults to perform a lexical retrieval task like persons with aphasia if the linguistic system’s processing efficiency is reduced in relevant ways. A few studies have induced aphasic-like performance in NBD participants on tasks of auditory comprehension, referential analysis, choice reaction time, and visual and auditory sequencing with relevant task demands (Miyake et al., 1994; Rosenthal & Bisiacchi, 1997; see also McNeil, 1982). In addition, aphasic-like processing of grammatical errors and auditory receptive agrammatic performance has been simulated in NBD adults (Blackwell & Bates, 1995; Dick et al., 2001). Although no simulations of lexical retrieval are yet reported in the literature, further insight into the underlying source and nature of the aphasic deficit will be provided if aphasic production deficits can be modeled in NBD individuals.

This study evaluated the picture naming performance of NBD young adults under time pressure and with competing cognitive tasks. Their performance was compared to that of NBD young adults on the same task without time stress. Error types from both groups were compared with those of persons with aphasia on a similar task. Temporal stress was manipulated in a picture-naming paradigm to attempt to simulate aphasic naming performance in the neurologically and cognitively intact adults. This parameter was chosen because of the belief that linguistic processing in aphasia is inefficient, as discussed above, and therefore slowed in individuals with aphasia, who are thus unable to process linguistic information as quickly as needed to complete the communicative task successfully. Using this logic, speeding the performance of NBD individuals was seen as one way to simulate this processing impairment. Competing cognitive tasks were also used to stress processing resources and attempt to increase the likelihood of error, as there is growing evidence in the literature that deficient attentional resource allocation is an integral part of aphasic deficits (e.g., Erickson, Goldinger, and LaPointe, 1996; McNeil et al., 1991; Murray, 2000; Murray, Holland, & Beeson, 1997; Petry, Crosson, Gonzalez Rothi, Bauer, & Schauer, 1994; Shuren, Smith Hammond, Maher, Rothi, & Heilman, 1995; Tseng et al., 1993). This factor was held constant across the experimental and control groups, so that the effect of temporal stress could be evaluated independently. It was hypothesized that NBD adults would produce more errors in a temporally stressed condition (the experimental group) than in a temporally unstressed condition (control group) and that the proportions of error types made by both of these groups would be similar to those made by persons with aphasia on a similar task without time stress. For this comparative purpose, error data were obtained from the authors of a previous study (Kohn & Goodglass, 1985), in which they recorded and categorized error responses of individuals with aphasia on a picture-naming assessment.

Method

Participants

Participants for this experiment were 60 college graduates between the ages of 20 and 30 years (30 men...
and 30 women) with no self-reported history of neurological damage or language impairment. Each participant scored within normal limits on the five-item version (Arvedson, McNeil, & West, 1986) of the Revised Token Test (RTT; McNeil & Prescott, 1978 [minimum score = 14.14]); on the CFL form of the Word Fluency Measure (WFM; Borkowski, Benton, & Spreen, 1967 [minimum score = 25]), on which they were asked to list words beginning with the letters C, F, and L; and on the Coloured Progressive Matrices (CPM; Raven, 1962 [minimum score = 22]), all tests that are sensitive to left and right hemispheric lesions (see Table 1). Seventy potential participants were screened, 4 of whom could not be included in the study because of a history of neurological injury or disease, and 6 of whom failed to meet criteria on the five-item screening version of the RTT. After stratification for gender, 30 participants were randomly assigned to each of the experimental and control groups.

Kohn and Goodglass (1985) described their participants with aphasia as follows:

Only those aphasics who correctly named at least 10 pictures on the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1976) were included in the present study. Forty-three male participants were tested, 39 of whom were right-handed. Each had suffered left-sided neurological damage, due to a cerebrovascular accident in 38 cases. Participants were classified by aphasia subgroup on the basis of both neurological criteria (CT scan lesion site and presence of hemiparesis) and linguistic criteria (Boston Diagnostic Aphasia Examination (BDAE), Goodglass & Kaplan, 1972). (p. 268)

Altogether, their participant pool included 9 participants classified as having Broca’s, 9 with Wernicke’s, 9 with conduction, 7 with frontal anomic, and 9 with posterior anomic aphasia. Age and severity of impairment were not specified. For this project, the authors were provided with error data from 34 of the participants from the Kohn and Goodglass (1985) data set, with no information provided about the diagnoses of those excluded.

### Stimuli

Picture stimuli were compiled from a variety of sources (Abbate & LaChappelle, 1984; Carlson, 1985; Dunn & Dunn, 1981; Kaplan, Goodglass, & Weintraub, 1983; Mayer-Johnson, 1981; McCarr, 1980), but were all black and white line drawings of picturable nouns, similar in style and visual complexity to the pictures on the Boston Naming Test (BNT). All portrayed words occurring less than five times per 1,000,000 words according to the American Heritage Word Frequency Book (Carroll, Davies, & Richman, 1971). Low frequency words were chosen to further stress lexical access and increase error productions. All stimuli were scanned into a MacIntosh II computer. A total of 80 pictures were used, arranged into 20 blocks of four pictures. Each block consisted of the four pictures presented simultaneously in a 2 × 2 array (see Figure 1). Ten of the blocks contained four semantically unrelated stimuli, and the remaining 10 blocks contained four stimuli of which two, three, or four were semantically related (5, 3, and 1 block of each, respectively). Semantic relatedness within stimulus blocks was manipulated to create the competing semantic decision task. The order of stimulus presentation was the same for all participants, and we controlled it for potential semantic priming effects between trial blocks by ensuring that there were no semantically related items presented in successive stimulus blocks. Phonological and prosodic priming effects were not controlled. We presented stimuli on the computer screen using a modified version of the Macintosh Reaction Time program (Chute & Daniel, 1988).

For the aphasic data, only those BNT items that were retained on a later version (Kaplan et al., 1983) were analyzed. This was done because the version of the BNT used by Goodglass and Kohn (1985) was not available to us for verification of the contents of the pictures, allowing no judgments to be made about certain visually based error types. In total, 61 items were analyzed for all participants with aphasia.

### Procedures

All control and experimental group participant responses were audiotaped for the purpose of linguistic analysis. Sessions were conducted individually in a quiet room.

Stimulus blocks were presented on a computer screen, with a short duration puretone (MacIntosh

<table>
<thead>
<tr>
<th>Group</th>
<th>Age</th>
<th>CPMa scores</th>
<th>RTTb scores</th>
<th>WFMc scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. score</td>
<td>24.8</td>
<td>34.8</td>
<td>14.59</td>
<td>46.8</td>
</tr>
<tr>
<td>SD</td>
<td>2.1</td>
<td>1.3</td>
<td>0.23</td>
<td>10.1</td>
</tr>
</tbody>
</table>

| Control | Avg. score | 24.3 | 34.7 | 14.64 | 46.2 |
| SD | 2.3 | 1.6 | 0.23 | 7.6 |

aColoured Progressive Matrices.  bRevised Token Test–5-item screening version.  cCFL form of Word Fluency Measure.

### Table 1. Summary of participant data.
“simple beep”) presented at a comfortable listening level simultaneously with picture presentation onset. Participants were instructed to name each of the pictures. They then stated whether any of the pictures had a semantic relationship and, if so, how many. For example, the format of the response for a set with pictures of a dog, computer, plate, and cat, in which the dog and the cat are considered semantically related, would be, “dog, computer, plate, cat, yes, two.” The participant was not required to designate which items were related or what the relationship was. When the participant had completed the entire response for a stimulus set, she or he pressed the space bar on the keyboard to initiate the next set. For the experimental group, this process was to be done as quickly as possible, as described below.

At least three practice trials, consisting of stimuli not in the experimental task, were presented before beginning the experimental task. This provided participants an opportunity to understand the response patterns for sets with and without semantically related stimuli and gave the experimenter a chance to give explicit feedback regarding the speed of the response. More than one presentation of these same practice stimulus blocks was permitted at the participants’ request following the prescribed three trials if there was uncertainty about the expected response pattern. In the experimental task, all participants in both the control and experimental groups were presented with two consecutive passes through the 20 sets of stimuli, in the same order both times, for a total of 40 blocks of stimuli, or 160 individual pictures. Two consecutive passes were used to increase the number of tokens collected from each participant.

Test procedures were identical between the experimental and control groups, with two exceptions. First, the experimental group was informed that response time was of critical importance during the test. They were encouraged to sacrifice accuracy if necessary for the sake of rapid response. Verbal encouragement to respond quickly (i.e., telling the participant, “faster”) was provided after every second to third stimulus block throughout the experimental task. Control participants were administered the same task with no time constraints. Further stress factors for both groups included the low frequency of occurrence of stimuli and the superimposition of an additional cognitive task (making decisions about the

Figure 1. Sample stimulus block as presented on the computer screen.
presence and, if present, number of the semantically related stimuli in order to increase competition for processing resources.

The second difference between groups was in the procedure used to verify the familiarity of target vocabulary and that the participant's target name for stimulus items was the same as the experimenter's. For this purpose, each participant in the experimental group was asked to view all stimuli for a third time, after the full experimental condition was completed, and to provide just the object names, with no time constraints or competing cognitive tasks. We conducted this verification with the control group by asking them about any item for which they did not produce the expected target name on either trial. A determination was made as to whether the incorrect response was, in fact, an error or merely reflective of vocabulary differences. If the target response was produced for one of the two trials with any given stimulus item for the control group, the nontarget response was considered an error. These different methods of target verification for the two groups were chosen on the assumption that control group participants had already gone through the list in an unhurried manner and thus did not need to repeat the entire list a third time. Experimental participants, however, had not had the chance to be sure they were providing the best name for an item, because of time constraints. For either group, any stimulus block that contained an item for which the participant's and the experimenter's target name did not agree was omitted from the analysis for that participant. All analyses were conducted on these reduced data.

The primary focus of this study was the effect of temporal stress on the nature of errors produced. Accordingly, the proportion of each error type observed within the control, experimental, and aphasic groups was analyzed. In addition, we compared the rates of response for the experimental and control groups to validate that temporal stress was, in fact, present for the experimental group.

**Intrajudge Reliability Measures**

Through the process of creating error code combinations as needed to adequately describe responses, we created 232 different error categories. Coding was performed by the first author. Responses that were particularly difficult to code were discussed with the second author until agreement was reached. Intrajudge reliability of the coding procedures was established by re-evaluating 5% of the coded responses randomly selected from each of the three participant groups (including the aphasic data). These responses reflected a range of error subtypes. For a recoding attempt to be considered the same as the original coding, it had to match exactly, including all components. For example, if the original coding was "semantic (in-class coordinate)/perceptual/negated" and the recoding yielded "semantic (in-class coordinate)/negated," this would be considered a discrepancy and count against the reliability of the coding procedure. Percent coding agreement was 83% for the control group responses, 86% for the experimental group, and 77% for the aphasic data. Overall intrajudge agreement (across all three groups) was 82%.

**Responses Eliminated From Analysis**

A total of 428 blocks (1,712 words) were omitted from analysis because of disagreement on target name in the experimental group, and 459 blocks (1,836 words) were omitted for this reason in the control group. In addition
to the responses omitted from analysis due to vocabulary discrepancies, some were omitted for time-related reasons. Response times were collected for each stimulus block. Response time was defined as the time between key presses, from the moment of presentation of a stimulus (designated by the key press which triggered the change in stimulus) until the next key press, which brought up the next stimulus set. For the experimental group, for whom response time was critical, a response that fell more than 2 standard deviations above the mean of the distribution of all experimental group responses was considered an outlier, indicative of potentially inadequate temporal stress, and the stimulus block was omitted from linguistic analysis, although included in temporal difference analysis. A total of 24 stimulus blocks (96 words) were omitted from analysis for this reason. An accurate response time was not obtained because of computer error for 74 stimulus blocks (296 words) in the experimental group and 49 stimulus blocks (196 words) in the control group. Because this information was used to determine that the two groups were different in terms of temporal stress, these stimulus blocks were excluded from the analysis for both groups, even though time was not a critical factor for the control group.

A total of 89 blocks (356 words) in the experimental group and 47 blocks (188 words) in the control group were omitted from analysis for some combination of vocabulary and time issues. In total, 2,580 responses from the control group and 2,340 responses from the experimental group were analyzed for this study, from a total of 4,800 responses obtained for each group. Aphasic errors were obtained from a total of 2,074 responses (61 items analyzed for each of 34 participants).

As discussed earlier, data were reduced to a single representation of each error type occurring for each target, regardless of how many attempts were made to name it. In this process, both the first and second presentations of each target for each participant were considered together. The total number of different targets analyzed was 1,352 for the control group and 1,324 for the experimental group. Analyses reported below were all conducted on this reduced data set. In addition, errors across all subtypes of uncodable responses, semantic errors, whole/part errors, and nonword errors were combined and represented in analyses as a single unit for each error type (i.e., all uncodable responses evaluated together, all whole/part errors evaluated together, etc.). This was done to simplify the analysis and interpretation, as results were not seen to change substantially when analysis was done for these very similar subtypes separately versus aggregating across subtypes. All analyses of group differences were completed on these 19 aggregated error categories. More detailed analysis of semantic error subtypes was also completed, as we discuss later.

### Results

#### Temporal Effects

A Mann–Whitney U test for unrelated samples indicated that the responses of the experimental group were significantly faster ($M = 5,866$ ms, $SD = 2,183$ ms), $U = 327,277, z = 22.45, p = .05$, than those of the control group ($M = 11,200$ ms, $SD = 5,496$ ms).

#### Group Performance

Twenty-six participants in the control group produced errors on 91 of the 1,352 targets analyzed for that group (6.7%) (4 participants in this group made no errors). All 30 participants in the experimental group contributed to errors made on a total of 380 out of 1,324 targets analyzed (28.7%). The 34 aphasic participants made errors on a total of 1,013 targets of the 2,074 targets analyzed in their task (48.8%), which were recoded for error type using the current coding system. All further analyses are based on the proportions of each error type relative to the number of targets in error. These proportions were calculated for each participant in the three groups, and were then used to compare the groups.

#### Between-Group Comparisons

A bootstrap procedure was used to determine 95% confidence intervals for the average proportion of each type of error made by each group (see Figure 2). Two sample t statistics were then calculated to compare the distributions of the proportion of each error type between each pair of participant groups (with the exception of those few error types for which neither group in the comparison produced any errors). For an overall comparison of each pair of groups within the three participant groups, the absolute values of the t statistics were then summed across all error categories. In a randomization test, we compared this overall score with the distribution of t scores obtained through repeatedly (10,000 times) assigning the participants at random to one of the two groups. The rank of the actually observed overall score relative to the simulated overall scores yielded a p value for the comparison of the two groups. All resulting pairwise comparisons of the three groups using all 19 aggregated error categories were found to be significant at $p = .0001$ (see Table 2), indicating that, overall, each of the groups was different from each of the other groups.

The majority of individual error types produced showed no difference in pairwise comparisons between groups. Any interpretation of data taken as support for the aphasia–normalcy continuum hypothesis, however, must take into account any significant differences found.
Therefore, results presented here describe those differences and their significance within the context of this hypothesis. Table 3 provides a breakdown showing all error types that yielded at least one significant (unadjusted, \( p \leq .05 \)) two-sample \( t \) statistic on pairwise comparison. Information about significance at this unadjusted level is provided to give perspective on the overall error patterns without extreme constraints. Given the many comparisons made in this process (56 total, as one error type was not produced by either the control or experimental groups, precluding inclusion in this analysis), however, we address below only those comparisons that were significant at a Bonferroni-adjusted significance level of .001.

Table 2. Results of randomization testing for comparison of groups across all error types.

<table>
<thead>
<tr>
<th>Group comparison across aggregated error categories</th>
<th>Actual summed ( t )</th>
<th>Range of summed ( t ) simulated in randomization test</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control vs. experimental</td>
<td>27.68</td>
<td>7.31–26.00</td>
<td>.0001*</td>
</tr>
<tr>
<td>Control vs. aphasic</td>
<td>51.53</td>
<td>6.65–34.50</td>
<td>.0001*</td>
</tr>
<tr>
<td>Experimental vs. aphasic</td>
<td>60.60</td>
<td>7.21–35.74</td>
<td>.0001*</td>
</tr>
</tbody>
</table>

*Significant at Bonferroni-adjusted significance level .0167.
As seen in Table 3, all groups differed from each other on the number of unrelated responses produced ($p \leq .0008$; see Table 3 for specific $t$ values and the direction of the differences discussed here). The control and experimental groups also differed in the proportion of multiword responses and rejections/denials ($p \leq .001$). The control and aphasic groups differed in the proportion of nonword approximations, neologisms, negations, rejections/denials, and related verbs produced ($p \leq .001$). The experimental and aphasic groups differed in the proportion of phonemic/phonetic errors, perceptual errors, whole/part errors, nonword approximations, neologisms, negations, multiword responses, and related verbs produced ($p \leq .001$).

Semantic errors (all subtypes considered together) accounted for more errors produced than any other single error type in all three groups (39.4% for the control group, 45.2% for the experimental group, and 22.2% for the aphasic group). Therefore, the distribution of semantic error subtypes was of interest. In the present investigation, the same method of bootstrapping and randomization testing to evaluate significant differences between groups was applied to the counts of semantic error subtypes relative to the total number of semantic errors. Using a Bonferroni-adjusted significance level of .017 for the overall comparisons, we found that the control and the experimental groups were not significantly different from each other but that both the control and experimental groups were significantly different from the aphasic group (see Table 4). However, the experimental and the aphasic groups were much more similar to each other than were the control and aphasic groups, and the overall pattern of semantic error subtypes produced was similar across all groups (Figure 3), with superordinate, in-class coordinate, and contextual associate errors collectively accounting for the vast majority of semantic errors for all groups (84.6% of semantic errors in the control group, 89.1% in the experimental group, and 89.2% in the aphasic participants). Analysis of $t$ statistics for specific semantic error

### Table 3. Two sample $t$ statistics and $p$ values from randomization tests for pairwise comparisons of all error types showing significant differences ($p \leq .05$, unadjusted) between groups.

<table>
<thead>
<tr>
<th>Error type</th>
<th>% of all errors&lt;br&gt;control; experimental; aphasic</th>
<th>Control vs. experimental&lt;br&gt;$t$</th>
<th>$p^b$</th>
<th>Control vs. aphasic&lt;br&gt;$t$</th>
<th>$p^b$</th>
<th>Experimental vs. aphasic&lt;br&gt;$t$</th>
<th>$p^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B – Phonemic/phonetic</td>
<td>16.0; 9.0; 24.7</td>
<td>1.44</td>
<td>.1569</td>
<td>-1.64</td>
<td>.1076</td>
<td>-4.60</td>
<td>.0002*</td>
</tr>
<tr>
<td>2C – Perceptual</td>
<td>14.2; 21.3; 7.9</td>
<td>-1.52</td>
<td>.1289</td>
<td>1.80</td>
<td>.0746</td>
<td>4.53</td>
<td>.0001*</td>
</tr>
<tr>
<td>2D – Whole/part</td>
<td>11.0; 17.2; 6.9</td>
<td>-1.14</td>
<td>.2640</td>
<td>0.95</td>
<td>.3819</td>
<td>3.62</td>
<td>.0005*</td>
</tr>
<tr>
<td>2E – Unrelated</td>
<td>0.0; 3.0; 15.9</td>
<td>-3.36</td>
<td>.0008*</td>
<td>-6.00</td>
<td>.0001*</td>
<td>-5.00</td>
<td>.0001*</td>
</tr>
<tr>
<td>2F – Nonword approximations</td>
<td>5.2; 1.2; 21.4</td>
<td>1.75</td>
<td>.0802</td>
<td>-4.30</td>
<td>.0001*</td>
<td>-6.75</td>
<td>.0001*</td>
</tr>
<tr>
<td>2G – Unrecognized correct</td>
<td>4.1; 0.6; 2.8</td>
<td>1.61</td>
<td>.1204</td>
<td>0.57</td>
<td>.5864</td>
<td>-2.24</td>
<td>.0217</td>
</tr>
<tr>
<td>2I – Neologism</td>
<td>0.0; 1.4; 25.8</td>
<td>-1.80</td>
<td>.1171</td>
<td>-6.60</td>
<td>.0001*</td>
<td>-6.60</td>
<td>.0001*</td>
</tr>
<tr>
<td>3 – Negation</td>
<td>1.5; 0.2; 17.4</td>
<td>0.94</td>
<td>.5921</td>
<td>-5.05</td>
<td>.0001*</td>
<td>-6.51</td>
<td>.0001*</td>
</tr>
<tr>
<td>4A – Adjective or adverb</td>
<td>0.0; 0.6; 1.1</td>
<td>-0.93</td>
<td>.7296</td>
<td>-2.96</td>
<td>.0026</td>
<td>-0.81</td>
<td>.4945</td>
</tr>
<tr>
<td>4C – Multiword responses</td>
<td>28.3; 6.3; 28.9</td>
<td>3.61</td>
<td>.0004*</td>
<td>-0.09</td>
<td>.9280</td>
<td>-4.93</td>
<td>.0001*</td>
</tr>
<tr>
<td>5 – Rejection/denial</td>
<td>16.6; 1.5; 0.1</td>
<td>3.15</td>
<td>.0010*</td>
<td>3.79</td>
<td>.0001*</td>
<td>1.42</td>
<td>.1598</td>
</tr>
<tr>
<td>7 – Nominalized description</td>
<td>1.3; 1.2; 3.2</td>
<td>0.06</td>
<td>.9989</td>
<td>-1.30</td>
<td>.2208</td>
<td>-1.97</td>
<td>.0458</td>
</tr>
<tr>
<td>8 – Related verb</td>
<td>0.0; 1.6; 4.8</td>
<td>-2.39</td>
<td>.0249</td>
<td>-6.64</td>
<td>.0001*</td>
<td>-3.49</td>
<td>.0012</td>
</tr>
</tbody>
</table>

*aPercentages of all errors are reported in the format “control; experimental; aphasic.”  

*bAll comparisons significant at unadjusted significance level = .05 are in bold print.

*Significant at Bonferroni-adjusted significance level .001.

### Table 4. Results of randomization testing for comparison of groups across semantic error subtypes.

<table>
<thead>
<tr>
<th>Group comparison across semantic error subtypes</th>
<th>Actual summed t</th>
<th>Range of summed</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>simulated in randomization test</td>
<td></td>
</tr>
<tr>
<td>Control vs. experimental</td>
<td>10.13</td>
<td>1.25–14.93</td>
<td>.0403</td>
</tr>
<tr>
<td>Control vs. aphasic</td>
<td>18.87</td>
<td>2.86–18.59</td>
<td>.0001*</td>
</tr>
<tr>
<td>Experimental vs. aphasic</td>
<td>14.30</td>
<td>2.52–18.01</td>
<td>.0127*</td>
</tr>
</tbody>
</table>

*Significant at Bonferroni-adjusted significance level .0167.
subtypes (Table 5) showed no significant differences between the control and experimental groups \( (p > .05) \), for all pairwise comparisons. At the Bonferroni-adjusted significance level of .002, the control and aphasic groups differed significantly only on the subtype of right noun/omitted or out-of-class adjective. The experimental and aphasic groups showed differences approaching significance only on correct adjective/omitted or out-of-class noun errors and subordinate errors \( (p = .004 \) and .008, respectively).

**Discussion**

This study was undertaken to compare picture naming errors for NBD participants under conditions of temporal stress and cognitive load versus under no temporal stress and versus persons with aphasia under no temporal stress or imposed cognitive load. We hypothesized that NBD young adults under stress of time limitations and competing cognitive tasks would be induced to produce naming errors more similar to those seen in aphasia than their non-time-stressed NBD counterparts. The results generally support this hypothesis.

Overall, placing NBD adults under time stress caused them to produce proportionally more errors than did NBD adults without time stress, although not as many as the participants with aphasia. In terms of the types of errors produced, the contrast of greatest interest is between the experimental and the aphasic groups. Among the two-way comparisons, this group contrast actually produced the largest number of significantly different proportions of error types. This appears to be contrary to the notion of time stress moving NBD adults closer to aphasic behavior, but a closer investigation of
Table 5. Two sample t statistics and p values from randomization tests for pairwise comparisons of all semantic error subtypes showing significant differences (p ≤ .05, unadjusted) between groups.

<table>
<thead>
<tr>
<th>Error type</th>
<th>Control vs. experimental</th>
<th>Control vs. aphasic</th>
<th>Experimental vs. aphasic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A(ii) – In-class coordinate</td>
<td>58.0; 46.0; 42.1</td>
<td>1.35</td>
<td>.1786</td>
</tr>
<tr>
<td>2A(iii) – Contextual associate</td>
<td>21.9; 29.5; 35.2</td>
<td>-1.29</td>
<td>.2052</td>
</tr>
<tr>
<td>2A(v) – Right noun, omitted or out of class adjective</td>
<td>0.0; 1.3; 2.4</td>
<td>-1.95</td>
<td>.0836</td>
</tr>
<tr>
<td>2A(vi) – Right adjective, omitted or out of class noun</td>
<td>1.9; 4.5; 1.1</td>
<td>-1.46</td>
<td>.1522</td>
</tr>
<tr>
<td>2A(vii) – Right adjective, in-class noun</td>
<td>7.6; 1.5; 1.5</td>
<td>1.89</td>
<td>.0696</td>
</tr>
<tr>
<td>2A(ix) – Subordinate</td>
<td>0.0; 0.0; 1.7</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Percentages of semantic errors are reported in the format “control; experimental; aphasic.” *All comparisons significant at unadjusted significance level = .05 are in bold. *NA because no control or experimental participant made an error of this type.

*Significant at Bonferroni-adjusted significance level .002.

The error types yielding differences indicates that, in fact, the experimental participants resembled the aphasic participants in many ways. Contrasts of the control group with each of the other groups provided a broader perspective on error patterns.

Some of the error types that yielded significant differences between the aphasic and experimental groups are likely the result of task differences between the groups rather than language processing differences induced by the tasks. In particular, the significantly greater proportion of perceptual and whole/part errors produced by the NBD individuals under time stress may be the result of increased stimulus complexity and reduced time available for perceptual processing by NBD participants. It is possible that the combination of time limitation and the need to process four pictures at once (given the format of stimulus presentation for this group) caused a decrease in the accuracy of visual analysis for the experimental group. The proportion of multiword responses and negations made by the experimental group—significantly less than the aphasic group—are also likely artifacts of the severe time limitations imposed on them preventing these kinds of longer responses. This is supported by the observation that the control group also produced significantly more multiword responses than the experimental group.

Two error types that appear to be true discriminators between the aphasic participants and either NBD group are nonword approximations and neologisms. These kinds of errors are commonly seen in aphasia and, even under severe time stress, they were not induced in the NBD adults. It is beyond the scope of this article to provide an exhaustive review of theories of neologistic speech, but there are some theories that can address the differences obtained here. Buckingham and Kertesz (1976), Butterworth (1979), Schwartz, Saffran, Bloch, and Dell (1994), and Shuren et al. (1995) all discussed possible mechanisms for neologistic language production that involve reduced self-monitoring as a contributory factor. Any of these explanations may account for the difference in neologistic productions seen between both NBD groups and the participants with aphasia in this study. The experimental group’s feedback and self-correction mechanisms may not have been sufficiently stressed by this task to simulate the breakdown of these systems observed in persons with aphasia, thus resulting in the actual production of few neologisms.

Another factor that may account for some of the group differences seen here is the use of compensations by the aphasic group. The presence of significantly more related verbs in the aphasic group data may be the byproduct of that group being more accustomed to failures in word retrieval and therefore better able to draw on circumlocution as a compensatory tool. The same may be true for the trend toward a greater proportion of adjective/adverb and nominalized description errors produced by the participants with aphasia, although these only approached significance at the stringent Bonferroni-adjusted criterion. This interpretation is supported by the finding that the control group was significantly more likely to reject a target than make an error, perhaps indicating that they were not as willing to use alternative or compensatory approaches when word retrieval failed. The trend toward greater use of related verbs by the experimental participants relative to control participants indicates an area in which experimental participants were moved away from the “normal” state and toward a state more similar to the group with aphasia.

This same kind of shift away from the control group and toward behavior consistent with the participants with aphasia was found for unrelated errors. It is unclear why the group with aphasia produced significantly
more unrelated errors than either of the NBD groups. It is possible that this finding indicates some kind of specific shift in the semantic system that was not induced adequately with increasing time pressure in the NBD adults. It is, however, interesting to note that the experimental group was moved significantly away from the control group and toward more aphasic-like proportions for this error type.

The final error type to be considered is the significantly greater proportion of phonemic/phonetically related errors produced by the aphasic group relative to the experimental group though, interestingly, not the controls. This seems to indicate that NBD adults under a condition of temporal stress were less susceptible to phonemic or phonetic impairment, which is contrary to previous reports in the literature (Dell, Burger & Svec, 1997; Fossett, McNeil & Pratt, 2001). One possibility is that the use of a semantically based competing cognitive task predisposed the experimental group participants to semantic errors. It would be important for future research to investigate whether this same pattern is found if a phonemically based competing cognitive task were used.

In terms of the distribution of semantic error subtypes, relative to all semantic errors, there is similarity despite the statistically significant differences between the aphasic group and both NBD groups. The overall pattern of semantic error subtypes produced by all groups is quite similar, with the same three subtypes accounting for the vast majority of semantic errors. This finding is consistent with the thesis that the basic structure of the underlying semantic system is unchanged by aphasia.

Overall, these results demonstrate that increased stress or processing competition can, in some instances, move nondisordered adults toward performance fundamentally similar to that demonstrated by individuals with aphasia. These results are consistent with findings by Schwartz et al. (1994), Miyake et al. (1994), and Murray (2000) and may, at least as a first pass, be consistent with a shared attentional mechanism as in a single channel bottleneck theory (Pashler, 1998) or with a resource allocation mechanism (McNeil et al., 1991; Navon & Miller, 2002).

There are additional factors to be considered in future investigations. The largest caveat to the interpretation of this study is that the difference in tasks between the NBD groups and the aphasic participants may have been the source of a number of the findings reported here. In addition, because different pictures were named, there was no control for several possible psycholinguistic effects such as imageability, word length, articulatory complexity, familiarity, and so forth (Nickels & Howard, 1995). Analysis of same-task data would be valuable for this reason. Future studies might also manipulate the type of competing cognitive task used to increase stress in the NBD participants to assess whether the semantic task biased participants toward making semantic errors. Likewise, the measurement of performance on both tasks in the dual task methods used in this study would be important to determine whether the imposed competing task actually imposed a sharing of processing resources or delayed concurrent processing in either of the tasks. Whether the age differences between the participants with aphasia and the other two groups disguised actual differences remains to be assessed. In spite of these limitations, however, this preliminary demonstration that many aphasic naming deficits can be simulated in NBD individuals under time and resource-limited conditions, as well as the continued similarities in patterns of error types seen between individuals with aphasia and nonstressed NBD individuals, is largely consistent with an impaired-access, single-channel bottleneck (Pashler & Johnston, 1989) and/or impaired resource allocation (McNeil et al., 1991) view of aphasia.

Conclusions

The results of this study generally support the idea of an aphasic-to-normal continuum of errors in linguistic performance (McNeil, 1982, 1988; Miyake et al., 1994; Miyake, Carpenter, & Just, 1995) and the notion of aphasia being largely the product of language processing impairments rather than fundamentally altered linguistic mechanisms (McNeil et al., 1991). In this study, we used both temporal stress and competing cognitive tasks to tax the normal language system, so further investigation of the effects of temporal stress alone may yield further insight into the relative contribution of time versus divided attention to the generation of these types of language errors.

Acknowledgments

This work was completed as a thesis in partial fulfillment of the requirements for an MS in Communicative Disorders at the University of Wisconsin–Madison. We are grateful to Harold Goodglass and Susan Kohn for their invaluable contribution of data. We thank Keith Kuhlmeier and the staff of the Trace Research and Development Center for their assistance with various aspects of the project. We are grateful to Gary Weismer and Margaret Lemme for their assistance and encouragement throughout the process and to Margaret Rogers for editorial assistance and constructive insight in preparation of the manuscript.

References


Appendix (p. 1 of 2). Error coding system.

**CATEGORY 1: Uncodable responses**

1A – When the participant said nothing.
1B – When the participant responded only with a general comment (“That’s a hard one,” “I have one at home”) or question (“Is it alive?” “What kind of animal is that?”).
1C – When the participant produced a broken-off noun phrase with a modifier (“a dog’s…” for muzzle).
1D – When the participant produced less than a syllable (“k…”).
1E – When the participant produced only a nonspecific name (e.g., “something,” “object,” “structure”).

**CATEGORY 2: Noun errors**

2A(i) – Semantic—superordinate (“plant” for flower).
2A(ii) – Semantic—in-class coordinate—a response which could be put into the same category as the target (“airplane” for helicopter, “checkers” for dominoes).
2A(iii) – Semantic—contextual associate—a response which is something you would expect to find in the same place as the target object (“desert” for camel, “easel” for palette).
2A(iv) – Semantic—materials from which the target object was made (“rope” for noose, “paper” for label).
2A(v) – Semantic—right noun, omitted or out-of-class adjective (“brush” for toothbrush). Use of this code required that the target word be made up of adjectival and nominal portions.
2A(vi) – Semantic—right adjective, omitted or out-of-class noun (“tooth” for toothbrush, “wheelform” for wheelchair). Use of this code required that the target word be made up of adjectival and nominal portions.
2A(vii) – Semantic—right adjective, in-class noun (“seafish” for seahorse). Use of this code required that the target word be made up of adjectival and nominal portions.
2A(viii) – Semantic—right noun, in-class adjective (“direction vane” for weathervane, “swimhorse” for seahorse). Use of this code required that the target word be made up of adjectival and nominal portions.
2A(ix) – Semantic—subordinate (“lily” for flower).
2A(x) – Semantic—right noun, unintelligible adjective (“clabrush” for toothbrush). Use of this code required that the target word be made up of adjectival and nominal portions.
2A(xi) – Semantic—right adjective, unintelligible noun (“toothplepah” for toothbrush). Use of this code required that the target word be made up of adjectival and nominal portions.
2A(xii) – Semantic—noun-adjective reversal (“brushtooth” for toothbrush). If this category was included in the coding of a particular response, all other codes applied to that response would regard the adjectival and nominal portions of the word in their reversed order (e.g., in the response “brushstall” for toothbrush, “brush” is regarded as a correct adjectival portion).
2B – Phonemic/phonetic errors—contained “significant” phonological elements of the target word, such as the
Appendix (p. 2 of 2). Error coding system.

initial consonant cluster (/bran/ for broom) or stressed syllable (/kevin/ for label). Errors that fell into this category could be either real words or jargon. If they were real words, they would be coded only in this category or possibly in a category of semantic or perceptual relation. If the given response was a neologism, it would be dual coded as a phonemically/phonologically related (2b) neologistic (2I) error.

2C – Perceptual errors—the participant named an object which was visually similar to the target (“windmill” for pinwheel, “lasso” for noose).

2D(i) – Whole/part (I)—the participant named a part of the target object (“eraser” for pencil).

2D(ii) – Whole/part (II)—the participant named a discrete physical entity which contained the target object (“clock” for pendulum, but not “desert” for camel).

2D(iii) – Whole/part (III)—the participant named something present in the target picture (“canoe” for paddle).

2E – Unrelated errors—the response was not semantically, perceptually, or phonologically related to the target word.

2F(i) – Nonwords (I)—the response was an approximation of the target word (“rhinosteros” for rhinoceros). To be coded as 2F(i), a response had to have no more than one consonant and one vowel different from the target word, and the response could not be a real word in itself.

2F(ii) – Nonwords (II)—the response was an approximation of a nontarget word (“hypnopotamus” for rhinoceros). To be coded as 2F(ii), a response had to have no more than one consonant and one vowel different from the word which was the subject’s target, and the response could not be a real word in itself.

2G – Unrecognized correct productions—when the participant continued naming the picture after producing the target word, or produced the target in a nontopical position (“You throw darts at it.”).

2H – Perseverations—the reutterance of a response (target or non-target) that had been used to name one of the previous 10 pictures. To be coded in this category, the response had to occur in exactly the same form in both responses (e.g., an early response of “sleeping” followed by a response to the next stimulus of “sleep” would not be considered a perseveration). Responses coded as perseverations were always dual coded, if only as unrelated responses.

2I – Neologisms—responses that were phonologically flawed to the point of not being clearly recognized as an attempt to produce an English word (/watf/ for raft, /domi:nos/ for dominoes, /rub:m/ for rhinoceros). These responses could, however, contain significant phonological elements of target words, and would then be dual coded with 2B.

CATEGORIES 3: Negated responses

“stapler, no,” “it’s not a stapler,” “paper clip, no,” for stapler. This category also included responses in which the participant gave a name followed by a sign of uncertainty such as “horse, but…” for unicorn. All negated responses were also coded by the response given along with the negation (e.g., “chair, no” for bed would be coded as both an in-class and contextual semantically related response as well as a negation).

CATEGORIES 4: Alternative/descriptive responses

4A – Single adjective or adverb, or adjective or adverb phrase (“dangerous” for darts)

4B – -ing forms of verbs in isolation (“swimming” for flippers, “sleeping” for cradle).

4C – Long noun phrases and other multiword responses— “to swim with,” “you wear it in the water” for flippers). This category included descriptive phrases (e.g., “A red one”) and onomatopoetic responses (“tweet tweet” for whistle), as well as a number of other multiword response forms. The general accuracy of these responses is captured in the other error codes assigned to them.

CATEGORIES 5: Rejection/failure to recognize/denial

“I don’t know,” “I don’t recognize that one.”

CATEGORIES 6: Nonpreferred correct

When the participant provided a form that could not technically be considered an error but which they later showed evidence of not using as their preferred name for that object (“postman” for mailman, “windvane” for weathervane).

CATEGORIES 7: Nominalized description

When the participant created a noun from a related verb (“blower” for whistle).

CATEGORIES 8: Related verb

“Sleep” for bed.

CATEGORIES 9: Spelling attempt

Spelling, or attempting to spell, both correct and incorrect responses, as well as responses such as, “Starts with an H.”