Analysis of Verbal Fluency Ability in Alzheimer’s Disease: The Role of Clustering, Switching and Semantic Proximities

Alyssa Weakley, Maureen Schmitter-Edgecombe*
Department of Psychology, Washington State University, Pullman, WA 99163, USA
*Corresponding author at: Department of Psychology, Washington State University, PO Box 644820, Pullman, WA 99164-4820, USA.
Tel.: +1-509-335-0170; fax: +1-509-335-5043
E-mail address: schmitter-e@wsu.edu (M. Schmitter-Edgecombe).
Accepted 26 February 2014

Abstract
The underlying nature of verbal fluency deficits in Alzheimer’s disease (AD) was investigated in this study. Participants were 48 individuals with AD and 48 cognitively healthy older adults. Fluency performance on letter and category tasks was analyzed across two 30-s intervals for total words produced, mean cluster size, and total switches. Compared with the control group, AD participants produced fewer words and switches on both fluency tasks and had a reduced category cluster size. The AD group was differentially impaired on category compared with letter fluency and produced more repetitive responses but fewer category exemplars than controls on the category task. A multidimensional scaling approach revealed that AD participants’ semantic maps were similar to controls. Overall, the data suggest that executive abilities involving search and retrieval processes and a reduced availability of semantically related words contributed to the AD group’s poorer performance despite similar temporal recall and organizational patterns.

Keywords: Alzheimer’s disease; Fluency; Language and language disorders; Executive functioning

Introduction
Cognitive deficits beyond memory impairment (e.g., speeded processing, executive functioning, and language difficulties) are common among individuals with Alzheimer’s disease (AD; Taler & Philips, 2008). Tasks of verbal fluency, which involve both language (temporal lobe) and executive functioning (frontal lobe) abilities, are consistently found to be impaired in AD populations in comparison with normative groups (Henry, Crawford, & Phillips, 2004). Despite considerable research, however, whether the underlying deficits on verbal fluency tasks reflect semantic network integrity, organization, strategy usage, and/or difficulties related to accessing and retrieving language material continues to remain unclear. The current study further explored the underlying nature of AD patients’ verbal fluency difficulties by examining the temporal production pattern and word retrieval strategies (i.e., clustering and switching) of AD patients completing letter (phonemic) and category (semantic) fluency tasks. Organization of the underlying language network was also explored in the category fluency task by measuring relationships between category exemplar responses.

Verbal fluency tasks evaluate an individual’s ability to rapidly retrieve and generate words under specific constraints including time (e.g., 60 or 90 s) and task rules (e.g., no repetitions, no names of people). Most neuropsychological evaluations include both category (e.g., names of animals) and letter (e.g., words that start with the letters F, A, and S) fluency tests, as they utilize different search strategies and tap into different functional areas of the brain. More specifically, category fluency involves both frontal and temporal lobe processes, as it requires a strategic search through semantic knowledge stores for words with associated characteristics (Birn et al., 2010; Mummery, Patterson, Hodges, & Wusem 1996; Taler & Phillips, 2008). As a result of this search, when “lion” is recited, the words “tiger” and “elephant” are more likely to be retrieved than “gopher” or “salmon” due to the close conceptual relation to animals found in the zoo. Letter fluency, on the other hand, primarily involves the frontal lobes (Birn et al., 2010; Mummery et al., 1996) and depends on knowledge of word spelling (e.g., words that start with the same letters) and phonemic
relatedness (e.g., homonym, words that rhyme) for word retrieval (Rohrer, Salmon, Wixted, & Paulson, 1999). Letter fluency also appears to tap the semantic network, although to a lesser extent than category fluency (Lezak et al., 2004).

Typical scoring procedures for verbal fluency tasks consist of tallying the total number of correct and error responses. Criteria have also been developed to assess clustering and switching strategies that are used to search, organize, and retrieve words (Troyer, Moscovitch, & Winocur, 1997). A cluster is a group of related words and relies heavily on semantic memory, knowledge of words (Murphy, Rich, & Troyer, 2006), and temporal lobe functioning (Troyer et al., 1997). Examples of category clusters include types of dogs (e.g., basenji, golden retriever) or animals found in Africa (e.g., lion, tiger, zebra). Letter fluency clusters include words that rhyme (e.g., truck, tuck), homonyms (e.g., aunt, ant), or words that begin with the same letters (e.g., sandwich, sand, sandal). A switch refers to a shift from one cluster of words to another and is associated with executive functions and the integrity of the frontal lobe to engage in flexible, strategic search processes (Troyer et al., 1997). Clustering has been considered more vital to category fluency performance, whereas switching is essential for both fluency tasks (Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998).

Although impaired on both fluency tasks compared with neurologically healthy controls, individuals with AD experience more prominent deficits on category compared with letter fluency tasks (Butters, Granholm, Salmon, Grant, & Wolfe, 1987; Haugrud, Crossley, & Vrbancic, 2011; Henry et al., 2004; Monsch et al., 1992; Murphy et al., 2006; Rascovsky, Salmon, Hansen, Thal, & Galasko, 2007; Troyer, Moscovitch, Winocur, Leach, & Freedman, 1998). In addition, studies have shown that category fluency declines at a faster rate than letter fluency in individuals with AD (Clark et al., 2009). Clustering and switching deficits have also consistently been found on category fluency in AD patients (Fagundo et al., 2008; Gocer March & Pattison, 2006; Haugrud et al., 2011; Murphy et al., 2006; Tröster et al., 1998; Troyer et al., 1998), and to a lesser extent on letter fluency (Beatty, Testa, English, & Winn, 1997; Troyer et al., 1998). Furthermore, the Boston Naming Test (BNT; Ivnik, Malec, Smith, Tangalos, & Petersen, 1996), a task that relies heavily on semantic memory, has been found to correlate more strongly with category than letter fluency (Henry et al., 2004; Weakley, Schmitter-Edgecombe, & Anderson, 2013). These findings suggest that in addition to impaired strategy usage, semantic network deficits may be negatively impacting AD performance.

Multidimensional scaling (MDS) is a statistical technique that has been used with category fluency tasks to examine semantic network organization. MDS visually constructs networks from participants’ verbal outputs by using Euclidean distances between pairs of produced category exemplars to establish a measure of similarity or dissimilarity. For example, the distance between cat and horse in the series “cat, dog, cow, horse” is three. The graphical representation of these distances provides the opportunity to establish dimensions of organization where the observed distances between points are assumed to reflect the proximity between words in an internal cognitive map. Studies that have examined MDS-derived maps suggest that a qualitative change in the semantic network of individuals with AD characterized by disrupted organization, weaker associations between closely related items (e.g., cat and dog), and uninterruptable and abnormal clustering of animals (Chan et al., 1993; Chang et al., 2013). A reduction in knowledge-based semantic characteristics (e.g., wildness—domesticity) has also been suggested (Chang et al., 2013).

Response patterns over a given fluency time interval provide additional, detailed information regarding organization, search, and retrieval strategies. Research suggests that word fluency comprises two factors: an initial semiautomatic retrieval phase, followed by a later effortful retrieval phase (Fernaeus & Almkvist, 1998; Ober et al., 1986). Therefore, a typical word fluency pattern is one of rapid word generation that tapers with task progression. This pattern also holds for switch and cluster production (Rabouet et al., 2010). Like healthy adults, both amnestic mild cognitive impairment (MCI; Weakley et al., 2013) and AD participants (Rohrer et al., 1999) have been found to produce a greater proportion of their responses in the earlier (semiautomatic) phase. Although some research suggests that AD participants differ in the rate of produced responses (defined by an earlier asymptote) from controls (Ober, Dronkers, Koss, Delis, & Friedland, 1986; Rosen, 1980), others have not found a significant decline in word production as a function of time (Lamar, Price, Davis, Kaplan, & Libon, 2002). To our knowledge, there has not been an examination of AD clustering and switching across time. To better understand organization strategies and word retrieval, in this study we examine word, cluster, and switch generation during two 30 s time intervals.

Previous verbal fluency research has suggested that individuals with AD experience a significant disruption of the semantic network. However, the mechanism underlying this disruption remains indistinct. Some researchers suggest that there is a substantive degradation of the semantic network in which knowledge representations are actually lost or significantly weakened (Chertkow & Bub, 1990; Henry et al., 2004). Others consider the semantic space to be disorganized (Chan et al., 1993; Chang et al., 2013). Still, some argue that semantic knowledge is preserved but a breakdown in executive control mechanisms responsible for strategy usage and/or effortful retrieval disrupts the ability to access semantic information (Bayles, Tomoeda, Kasznik, & Tosset, 1991; Grande, McGlinchey-Berroth, Milberg, & D’Esposito, 1996; Nebes, Brady, & Huff, 1989).

In the current study, we comprehensively evaluated the verbal fluency performances of AD patients to provide detailed information about the underlying nature of AD patients’ verbal fluency difficulties. We analyzed category and letter fluency performances using traditional scoring methods (i.e., total words, set-loss errors, repetitions) and methods sensitive to frontal and temporal lobe contributions (i.e., clustering and switching) across two 30 s time intervals. In addition, we examined the
organization of words recalled for category fluency using an MDS approach. A Procrustes transformation on the MDS coordinates was also employed to determine if meaningful differences between the cognitive maps remained after the data points were rotated, scaled, and/or reflected. Previous studies have independently examined these factors (with the exception of Procrustes transformation); yet, no study has combined all of the methods in a single study to determine how the different pieces of information fit together. Based on previous research, we hypothesized that AD participants will be impaired on both letter and category fluency tasks in comparison with controls, with more pronounced impairment on category fluency. In terms of clustering and switching abilities, both are expected to be impaired. Although reduced compared with controls, AD participants should show a similar pattern of word, switch, and cluster retrieval across time as controls. Finally, as has been shown in the previous literature (e.g., Chan et al., 1993; Chang et al., 2013), we hypothesized that our AD group would demonstrate disorganized MDS maps in comparison with controls. Overall, we anticipated that the data would show AD participants to have strategic word retrieval difficulties, disorganized word retrieval, and reduced semantic network integrity.

Method

Participants

Participants were 48 persons with AD (20 female, 28 male) matched on age and education with 48 cognitively healthy older adults (35 female, 13 male; see Table 1). All participants were age 58 years or older and able to provide informed consent or assent. Participants were volunteers tested as part of two larger researcher projects at Washington State University (see Schmitter-Edgecombe, Parsey, & Cook, 2011; Schmitter-Edgecombe, Woo, & Greeley, 2009). Both studies were reviewed and approved by the Washington State University Institutional Review Board.

Participants were recruited through advertisements, community health and wellness fairs, referrals from local agencies and physicians, and past studies. Exclusion criteria for AD and control participants included history of significant head trauma, current or recent (past year) psychoactive substance abuse, history of cerebrovascular accidents, and medical, neurological, or psychiatric causes of cognitive dysfunction. Participants who met initial screening criteria completed a 3-h battery of standardized and experimental neuropsychological tests. Each participant appointed a knowledgeable informant (e.g., spouse, adult child) who was contacted to supply subjective information about functional and cognitive ability and completed the Clinical Dementia Rating (CDR) Scale (Morris, 1993), which was administered by a CDR certified examiner. Medical information was also reviewed when available. All participants were given a report reviewing their performance as compensation for their time.

Participants were considered to have AD if they met research criteria of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS–ADRDA; McKhann et al., 1984) and were free of severe depression as documented by a score <10 on the 15-item Geriatric Depression Scale (Yesavage et al., 1983). On the CDR, 31% of the AD participants received a score of 0.5 (very mild dementia), 52% scored a 1.0 (mild dementia), and 17% scored a 2.0 (moderate dementia). The mean score on the Telephone Interview for

Table 1. Table of demographic and test variables

<table>
<thead>
<tr>
<th>Variable or test</th>
<th>AD (n = 48)</th>
<th>Control (n = 48)</th>
<th>t</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>75.10</td>
<td>74.40</td>
<td>.43</td>
<td>.09</td>
</tr>
<tr>
<td>Education (years)</td>
<td>15.52</td>
<td>15.89</td>
<td>.64</td>
<td>.13</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>42</td>
<td>73</td>
<td>.43</td>
<td>.25</td>
</tr>
<tr>
<td>TICS total score</td>
<td>24.29</td>
<td>34.48</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>Neuropsychological correlates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BNT total correct</td>
<td>45.07</td>
<td>56.42</td>
<td>.62</td>
<td>.18</td>
</tr>
<tr>
<td>Trails B (s)</td>
<td>239.47</td>
<td>85.10</td>
<td>.67</td>
<td>.50</td>
</tr>
<tr>
<td>Clox 1 total score</td>
<td>8.76</td>
<td>12.76</td>
<td>.58</td>
<td>.38</td>
</tr>
<tr>
<td>Dfswitch total score</td>
<td>2.00</td>
<td>7.58</td>
<td>.80</td>
<td>.23</td>
</tr>
</tbody>
</table>

Note: Unless otherwise indicated, mean scores are raw scores. Norm sources for the cognitive tests are in parentheses following the test. TICS = telephone interview of Cognitive status (Brandt & Folstein 2003); BNT = boston naming test (Ivnik et al. 1996); Trail making test, trails B (Reitan, 1958); Clox 1 (Royall et al. 1988); Dfswitch = switching subtest a component of the Delis–Kaplin executive functioning system design fluency subtest (Delis, Kaplan, & Kramer, 2001). *p < .001.
Cognitive Status (Brandt & Folstein, 2003) was a 24.29 (range 6–34). Participants classified as cognitively healthy older adults met the following criteria: (i) no self or informant reported history of cognitive changes, (ii) a CDR of 0, (iii) did not meet criteria for MCI, and (iv) no severe depression.

Procedure

All participants were administered the Delis–Kaplan Executive Functioning System (D-KEFS; Delis et al., 2001) letter and category verbal fluency subtests, a component of the D-KEFS Verbal Fluency subtest, along with a battery of other neuropsychological measures. The letter fluency subtest required participants to name as many words as possible starting with the letters F, A, and S for a total of 60 s each. Participants were instructed to refrain from saying names of people, places, numbers, as well as words that ended with different suffixes (e.g., run, runs, running). The category fluency subtest required participants to name as many animals as possible in 60 s.

Three raw scores were obtained for the letter (F, A, and S) and category (animal) verbal fluency subtests: total responses, average cluster size, and total switches. Scores were calculated for two intervals: 1–30 s (Interval 1) and 31–60 s (Interval 2). Total responses were scored by tallying all of the words recited excluding set-loss errors (e.g., names of people, places) and repetitions. The method derived by Troyer and colleagues (1997) was followed to score clustering and switching. If a cluster straddled between Interval 1 and Interval 2 the cluster was applied to whichever interval more words fell in. For example, if the words “lion, tiger, zebra” fell within Interval 1 and “monkey, elephant” fell in Interval 2, the entire cluster of words would be applied to Interval 1 only. If there was an equal number of words that fell in each interval, the cluster and switch was applied to Interval 2. Modifications to the Troyer (2000) scoring categories were made to include items specific to the Pacific Northwest region where testing took place (e.g., coyote under North American animal, llama under farm animal). All protocols were scored blind to participant diagnostic category by two independent raters. Repetitions and set-loss errors were also examined across both fluency tasks.

MDS was used to assess AD and control groups’ organization of words produced on the category fluency task. MDS is a mathematical tool used to analyze and graphically display proximity data to indicate the degree of dissimilarity between two things (Young & Harris, 1993). In the current study, animal names ranked as more similar were graphically generated in closer temporal proximity, whereas dissimilar items were more widely spread out (Moelter et al., 2001).

Prior to running MDS, 12 target words were selected based on how frequently they were produced by both participant groups. The most common words, in descending order, were: dog, cat, horse, cow, elephant, lion, tiger, deer, pig, sheep, zebra, and squirrel. Next, each participant’s response pattern of the selected 12 words was transformed onto a 12 × 12 matrix to depict the proximity data, with each cell representing the distance between two animals. For example, if a participant recited “dog, cat, pig, peacock, elephant” the cell that represents the distance between dog and cat would be entered as 1, and the distance between cat and elephant would be 3. Higher numbers represent less temporal proximity (dissimilarity) between animal names. Although errors were excluded from analyses, errors contributed to the proximity score as they occupied a distance integer.

Participants did not consistently name all words on the target list, leaving the proximity matrices of participants with too many missing values to be analyzed individually. Participant data were, therefore, combined by group (i.e., AD, control) using the method described by Chan and colleagues (1993), Paulsen and colleagues (1996), and Rossell, Rabe-Hesketh, Shapleske, and David (1999). The formula calculates the distance between each animal pair in the matrix (\(D_{ij}\)) between target words \(i\) and \(j\) of all subjects \(k\), controlling for the number of total responses that each participant made, \(n_k\). Group is identified as \(T\) (refer to Chan and colleagues (1993) or Paulsen and colleagues (1996) for an illustrative calculation example). As a result of the formula used, one combined proximity matrix was formed for the AD participants and another for the control group.

\[
D_{ij} = \frac{N}{(T_k)} \sum_{k=1}^{T_k} d_{ij/n_k}
\]

Analysis

Prior to running analyses, frequencies were investigated for letter and category fluency raw scores for total responses, mean cluster size, and total number of switches. Participants who performed 3 \(SD\) above or below the mean performance on any of the dependent measures were removed from all analyses for that particular fluency subtest. For example, if an individual performed 3 \(SD\) above the mean on total switches during the first 30 s of the letter fluency task they were removed from all analyses involving letter fluency, but were retained for the category fluency analyses. Three (1 AD, 2 control) participants were removed as outliers from letter fluency analyses and four (1 AD, 3 control) were removed from category fluency analyses. Four (3 AD, 1 control)
participants who did not complete the category fluency task were also removed from category fluency analyses. No participants were removed as outliers from both fluency tasks. Participant’s who did not complete the animal fluency task (3 AD, 1 control) or were unable to generate any of the 12 target responses (2 AD) were not included in the MDS analysis.

The dependent measures (i.e., total responses, average cluster size, number of switches) were analyzed using a group (i.e., AD, neurologically healthy older adults) by interval (i.e., Interval 1, Interval 2) mixed model analysis of variance (ANOVA) with repeated measures on the second factor for each fluency task (i.e., letter, category). Because \( \chi^2 \) analysis indicated a significant difference in gender between the AD and control groups, \( \chi^2 = (1, N = 96) = 9.58, p = .002 \), all analyses were also run with gender as a covariate. Since changes in findings were not observed, analyses without the covariate are reported.

Paired-samples \( t \)-tests were calculated for each participant group to determine if differences existed between participant groups’ Interval 1 response, cluster, or switch scores and their Interval 2 scores. Repetitions, set losses, and frequently produced words (high-frequency exemplars extracted for MDS analysis) for both groups were also submitted to \( t \)-tests as these measures can provide information regarding integrity of the associated brain areas. To examine differences between total words, switches, and mean cluster size produced on each fluency task, \( z \)-scores were calculated using the means and \( SD \) of the control group. Paired-samples \( t \)-tests were then performed to compare participant groups’ letter to category scores.

Correlations were performed between the fluency tasks and neuropsychological tests of language (i.e., BNT; Ivnik et al., 1996) and executive functioning (i.e., Clox1; Royall, Cordes, & Polk, 1998; Trails B; Reitan, 1958; switching subtest of D-KEFS Design Fluency subtest; Delis et al., 2001) that research suggests play a role in verbal fluency performance (Taler & Phillips, 2008). A significance level of \( p < .01 \) was used for Pearson correlations.

Finally, MDS was performed in SPSS using the PROXSCAL algorithm. While previous research has utilized ALSCAL (Chan et al., 1993; Chang et al., 2013; Paulsen et al., 1996), ALSCAL minimizes the squared Stress (S-Stress), giving too much “weight” to the extreme items in the stimulus configuration (Borg, Groenen, & Mair, 2013). PROXSCAL, a counterpart of ALSCAL, minimizes raw normalized stress resulting in a more accurate Euclidean space, therefore, this approach was utilized. Because iterative MDS algorithms are more flexible than classical MDS, and to optimize the MDS solution, starting configuration was randomly set 10,000 times. Random starts compute \( N \) MDS solutions, each beginning with a different configuration determined by the program until the solution with the lowest Stress is determined. Strict iteration criteria were set with stress convergence and minimum stress set to 0.000001 and the maximum iteration to 1,000 as recommended by Borg and colleagues (2013).

Goodness of fit was assessed by Stress-1 or “Kruskal’s Stress” level accounted for by the model. A Stress level of 0 indicates a perfect model fit. If the resultant map adequately represents the distance between animal names, then the stress levels should be low. Selection of a dimension for interpretation was made by choosing the solution that accounted for the least amount of Stress while remaining interpretable.

To compare individual group’s MDS solutions a Procrustes transformation (Gower & Dijkstra, 2004) was performed in MatLab version R2007b. The control group’s plot configuration was selected as the fixed target while the AD group’s configurations were optimally fitted to the target. Differences that remained after a Procrustes fitting were considered significantly meaningful and interpretable (Borg & Groenen, 2005).

Results

Neuropsychological Tests

Several of the neuropsychological tasks administered were similar across the two test samples from which the participants were drawn from including BNT, Trail Making Test, Design Fluency, and Clox. The AD group performed significantly more poorly than the control group on each of these measures (see Table 1).

Letter Fluency

Total responses. Mean scores and SD for each dependent measure can be found in Table 2. A 2 (group) \( \times \) 2 (interval) mixed model ANOVA conducted on total words revealed that more responses were generated at Interval 1 (\( M = 19.19 \)) than Interval 2 (\( M = 11.35 \)), \( F(1, 94) = 166.34, MSE = 17.07, p < .001, \eta^2 = .64, \) and by the control group (\( M = 37.73 \)) compared with the AD group (\( M = 23.42 \)), \( F(1, 94) = 21.83, MSE = 112.30, p < .001, \eta^2 = .19. \) The two-way interaction was not significant, \( F = .52. \) Paired-samples \( t \)-tests also showed a significant decrease in word production from Interval 1 to Interval 2 for both groups (\( t \)'s > 7.79; \( p \)'s < .001).

Cluster size. A mixed model ANOVA lacked significance for group (\( F = 2.46, \) interval (\( F = 1.46, \) and interval by group interaction (\( F = .00). \) Differences in size of clusters were not observed between Interval 1 and Interval 2 for either group, \( t \)'s > .80.
Switching. The ANOVA conducted on total switching revealed that more switches were made at Interval 1 (M = 12.38) than Interval 2 (M = 10.06), F(1, 94) = 21.04, MSE = 12.20, p < .001, η² = .18, and by the control group (M = 26.10) compared with the AD group (M = 18.83), F(1, 170) = 6.97, MSE = 35.02, p < .001, η² = .11. A two-way interaction was not significant (F = .63). Paired-samples t-tests showed a significant decrease in switch production from Interval 1 to Interval 2 for both groups (t’s = 2.88; p’s ≤ .006).

Category Fluency

Total responses. Mean scores and SD can be found in Table 3. A group by interval mixed model ANOVA conducted using total responses on category fluency revealed that more responses were made at Interval 1 (M = 10.98) than Interval 2 (M = 4.93), F(1, 86) = 284.03, MSE = 5.66, p < .001, η² = .40, and by the control group (M = 19.08) compared with the AD group (M = 10.46), F(1, 86) = 56.49, MSE = 5.66, p < .001, η² = .57. A group by interval interaction was not significant, F = 1.93. Paired-samples t-test revealed a significant decrease in word production from Interval 1 to Interval 2 for both groups (t’s > 2.88; p’s ≤ .006).

Cluster size. The main effect for interval (F = 1.28) and group by interval interaction (F = .36) lacked significance. A significant main effect for group revealed that the control participants produced larger mean cluster sizes (M = 1.16) than AD group participants (M = .87), F(1, 86) = 9.85, MSE = 0.61, p = .002, η² = .10. Paired-samples t-tests indicated that cluster size did not differ significantly between Interval 1 and Interval 2 for either participant group (t’s > .34).

Switching. An ANOVA conducted using total switches on category fluency revealed that more switches were generated at Interval 1 (M = 4.88) than Interval 2 (M = 3.03), F(1, 86) = 58.68, MSE = 2.54, p < .001, η² = .41, and by the control group (M = 9.38) compared with the AD group (M = 6.17), F(1, 86) = 18.71, MSE = 5.78, p < .001, η² = .00. A group by time interaction was not significant (F = .14). Paired-samples t-tests showed a significant decrease in switching ability from Interval 1 to Interval 2 for both groups (t’s > 5.25; p’s < .001).

### Table 2. Performances on the letter fluency task by group and time interval

<table>
<thead>
<tr>
<th></th>
<th>Total responses</th>
<th>Mean cluster size</th>
<th>Total switches</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 1</td>
<td>15.40 (8.87)**</td>
<td>.27 (.43)</td>
<td>10.35 (6.72)***</td>
</tr>
<tr>
<td>Interval 2</td>
<td>8.00 (5.51)**</td>
<td>.21 (.32)</td>
<td>8.46 (5.56)**</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 1</td>
<td>22.98 (9.23)***</td>
<td>.39 (.58)</td>
<td>14.40 (6.74)</td>
</tr>
<tr>
<td>Interval 2</td>
<td>14.71 (8.10)</td>
<td>.33 (.41)</td>
<td>11.67 (5.51)</td>
</tr>
</tbody>
</table>

Note: Mean scores are raw scores, SDs are in parentheses; Interval 1 = 0–30 s; Interval 2 = 31–60 s.
* p < .05.
** p < .01.
***p < .001 difference from control group.

### Table 3. Performances on the category fluency test by group and time interval

<table>
<thead>
<tr>
<th></th>
<th>Total responses</th>
<th>Mean cluster size</th>
<th>Total switches</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 1</td>
<td>8.52 (3.42)**</td>
<td>.85 (.53)*</td>
<td>4.05 (2.45)***</td>
</tr>
<tr>
<td>Interval 2</td>
<td>2.98 (2.39)**</td>
<td>.68 (.76)*</td>
<td>2.30 (1.69)***</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interval 1</td>
<td>13.43 (3.32)***</td>
<td>1.16 (.60)</td>
<td>5.70 (2.03)</td>
</tr>
<tr>
<td>Interval 2</td>
<td>6.89 (3.64)</td>
<td>1.10 (.93)</td>
<td>3.77 (1.92)</td>
</tr>
</tbody>
</table>

Note: Mean scores are raw scores, SDs are in parentheses; Interval 1 = 0–30 s; Interval 2 = 31–60 s.
* p < .05.
** p < .01.
***p < .001 difference from control.
Performance Differences Between Tasks

Paired-samples t-tests revealed that the AD group’s total word production was significantly more impaired on the category fluency task \( (M = -1.55) \) compared with the letter fluency task \( (M = -1.09); t(1, 42) = 3.87, p = .001 \). Total switches and mean cluster size scores, on the other hand, did not significantly differ by fluency task for either group (\( t \)'s between 2.30 and .33).

Examination of repetitions and set losses revealed that AD participant’s \( (M = 1.09) \) made significantly more repetitions on the category fluency task compared with controls \( (M = .40); t(1, 89) = 3.45, p = .001 \). Letter fluency repetitions and set losses and category fluency set losses were similar for both groups (\( t \)'s between −1.11 and 1.89).

Correlates with Neuropsychological Variables

Correlations revealed that all three of the executive functioning measures (i.e., Trails B, Design Fluency switching, Clox1) were significantly related to total responses on letter fluency for the AD group \( (r \)'s < −.49 or > .34; see Table 4). AD BNT score, on the other hand, was significantly related to total words and switches made on category fluency \( (r = .55 \text{ and } .51, \text{ respectively}) \). For the control group, the BNT total correct score correlated with category fluency total words \( (r = .39) \) and letter fluency switches \( (r = .42) \).

Multidimensional Scaling

Prior to applying MDS, the proportion of the 12 most frequently produced words, referred to as high-frequency exemplars, to total semantic words was compared across groups. While the control group \( (M = 6.35) \) produced significantly more high-frequency exemplars than the AD group \( (M = 5.00); t(94) = -2.42, p = .02 \), the proportion of high-frequency exemplars to total words for the AD group \( (M = .58) \) was larger than controls \( (M = .09); t(93) = -6.13, p = .02 \). Furthermore, the AD group made more high-frequency \( (M = 1.56) \) than low-frequency exemplar repetitions \( (M = .67; t(22) = 3.54, p = .002) \), and made more high-frequency exemplar repetitions in the first 30 s \( (M = .63) \) of the task compared with the control group \( (M = .17; t(33.42) = 2.09, p = .04) \).

For the control group, when all 12 variables were included in a 2D analysis, the Stress level (0.24) was the same as would be expected if the data were random \( (Spence & Ogilvie, 1973) \). Examination of the decomposition of normalized raw Stress by animal revealed an outlier for the word squirrel (0.08). This word was removed because it contributed disproportionately to the total Stress, it did not appear to form an intuitive cluster with any of the 11 remaining words, and it was the least commonly produced of the 12 target words. When the analysis was re-ran the Stress level dropped significantly to 0.15 which is well–below random Stress \( (Stenson & Knoll, 1969) \). Only a 2D solution was considered as it met Kruskal and Wish’s \( (1978) \) criteria regarding dimensionality, specifying that the number of stimuli (11 in the present case) should be at least four times greater than or equal to the dimensionality. A 3D solution falls beyond this criteria and is, therefore, not considered as it would result in an overfitting of the data.

Table 4. Correlations between neuropsychological tests and fluency task by group

<table>
<thead>
<tr>
<th></th>
<th>AD</th>
<th></th>
<th>Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total words</td>
<td>CL</td>
<td>Total switches</td>
<td></td>
</tr>
<tr>
<td><strong>Letter fluency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clox1</td>
<td>0.31*</td>
<td>0.26</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>DFswitch</td>
<td>0.34*</td>
<td>0.27</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Trails B</td>
<td>-0.49*</td>
<td>-0.33</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>BNT</td>
<td>0.36</td>
<td>0.23</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td><strong>Category fluency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clox1</td>
<td>0.35</td>
<td>0.22</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>DFswitch</td>
<td>0.34</td>
<td>-0.07</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Trails B</td>
<td>-0.39</td>
<td>0.03</td>
<td>-0.32</td>
<td></td>
</tr>
<tr>
<td>BNT</td>
<td>0.55**</td>
<td>0.14</td>
<td>0.51**</td>
<td></td>
</tr>
</tbody>
</table>

Note: Scores are correlation values represented by \( r \); DFswitch = switching subtest of design fluency; BNT = Boston Naming Test.

\* \( p < .01 \).
\** \( p < .001 \).
The MDS solution for the control group yielded two defined primary axes (see Fig. 1). Primary Axis 1 appears to be related to *exoticism* with animals found in the zoo (e.g., zebra, elephant) lying at one end of the spectrum and animals that are more common (e.g., cat, cow, pig) at the other. Of note, animals that could be considered pets fell at the midline. Deer fell at the extreme end of the common animal dimension, which may not be surprising given the location of the university where the data were collected (i.e., Pacific Northwest region). Primary Axis 2 appears to be related to *hoovedness*. That is, animals with hooves fell at one end of the semantic map while animals with paws/pads at the other end. Again, deer fell at the extreme, suggesting that an undefined attribute was assigned to this animal or participants were clustering it with animals not chosen for the MDS analysis. In addition to the two
primary axes, three “neighborhoods” appear in the plot. More specifically, when a vertical line is drawn down the middle of the plot a “zoo animal” neighborhood is discerned on the left side. At the top right the words cat and dog seem to cluster together, whereas the bottom right quadrant appears to be reserved for “farm animals.” It is also worth noting the proximity of lion and tiger to each other based on the number of similar traits (e.g., zoo, feline, pawed, Sahara desert) along with horse-cow (farm, hooved, large), and cat-dog (pet, domestic, small).

Fig. 3. AD and control group combined MDS common space plots.

Fig. 4. Procrustes transformed MDS coordinates.
When the AD group’s MDS solution contained all 12 words the resulting stress value also equaled 0.24. To keep plots comparable, and to reduce Stress, squirrel was removed. The result was a reduction in stress to 0.18, which is below the acceptable stress level (Stenson & Knoll, 1969). Similar to the control group’s solution, primary Axis 1 appeared to be related to exoticism with common animals falling at one end and zoo animals falling at the other (see Fig. 2). Like the control group’s map, animals that could be considered pets fell around the middle of the plot. A second axis was not detected. Neighborhoods also appeared to exist on the AD’s semantic map, although they appeared to be less well defined.

To determine if meaningful and interpretable differences existed between the groups’ MDS solutions Procrustes transformation was employed. Fig. 3 represents the maps’ original coordinates superimposed onto each other. Fig. 4 represents the transformed coordinates. The transformation resulted in the AD data being reflected around the Y-axis 180°, rotated ~−20°, and shrunk 1.6 times (scaled). A translation scalar of 0 indicates that the AD points were not moved left/right or up/down in the x-y plane. Once the AD points were flipped, rotated, and then scaled the AD coordinates appear to roughly align with control coordinates. Based on these results, AD participants are suggested to have very similar semantic maps in terms of organization of commonly retrieved words on the category fluency test.

Discussion

Although individuals with AD are known to experience impaired verbal fluency performance, whether the deficits are due to a breakdown in the integrity of their semantic network, a disorganization of conceptual knowledge, an impaired ability to access or retrieve information, and/or a deficit in strategic task performance remains unclear. The purpose of this research was to examine these potential factors, which may underlie the impaired verbal fluency test performance of individuals with AD, in a single study.

Results from the present study are consistent with the notion that individuals with AD experience pathological changes in both the frontal and temporal lobe structures as a result of the disease process. Frontal lobe impairment in AD was suggested by fewer words and switches between clusters of words on both fluency tasks, implicating poorer executive functions. Neuropsychological measures related to executive functioning (e.g., Clox 1, Trails B, D-KEFS Design Fluency) were also correlated with the AD groups’ letter fluency total responses. Analogous to previous research (Butters et al., 1987; Haugrud et al., 2011; Henry et al., 2004; Monsch et al., 1992; Murphy et al., 2006; Rascovsky et al., 2007; Troyer et al., 1998), total word production was significantly more impaired on category than letter fluency suggesting impaired temporal lobe involvement. In addition, compared with controls, mean cluster size was impaired on category but not letter fluency suggesting a distortion of the semantic network rather than a deficit in the ability to strategically make clusters. Temporal lobe degradation was further suggested by positive correlations between the AD group’s BNT performance, a task thought to rely heavily on semantic network integrity, and total words and switches produced on category fluency.

Despite the poorer word fluency performances of the AD participants, similar temporal patterning was found for word, cluster and switch production suggesting that individuals with AD approached the verbal fluency tasks in a similar manner to healthy older adults. Consistent with time-dependent attentional processes (e.g., Fernaeus & Almkvist, 1998) and Lamar and colleagues findings (2002), both the AD and control groups generated significantly more words and switches during the first 30-s interval of the letter and category fluency tasks compared with the second 30-s interval. The difference in switch and word production from Interval 1 to Interval 2 was also similar for both participant groups. In addition to suggesting that the AD and control groups approached the verbal fluency tasks in a similar manner, these results further indicate that the AD participants had the capacity to sustain mental set across the 60 s fluency tasks.

The MDS analysis, post Procrustes transformation did not support the notion of a disorganized semantic network in AD as had been previously suggested (Chan et al., 1993; Chang et al., 2013), at least for more common animal names produced by mild-to-moderate AD patients. Rather, after Procrustes transformation was applied, the AD group produced a remarkably similar semantic map as the control group (see Fig. 4). The control group’s MDS solution resulted in two recognizable primary axes (i.e., exoticism on horizontal axis and hoovedness on vertical axis), as well as animals with more traits in common being grouped closer together (e.g., lion and tiger, cat and dog, horse, and cow). Prior to performing Procrustes transformation, similar to Chan and colleagues (1993) and Chang and colleagues (2013), we could have concluded that the AD group had a more disturbed semantic network, with limited organizational strategies for word retrieval. Yet, as observed after applying a Procrustes transformation (see Fig. 4), the data suggest that the AD group did, in fact, rely on similar organizational strategies as controls when producing common category exemplars for the animal fluency task.

These results inform us of the importance of considering how cognitive maps are derived by MDS. In the current study, the MDS solution was optimized by trying 10,000 different starting configurations with 1,000 iterations to result in the most accurate solution with lowest Stress. Because there could be a number of equally good solutions selected by the model it is not improbable for two different clinical groups to end with MDS solutions that are seemingly different, but interpretable similar. After
Procrustes transformation, any differences found between two or more MDS solutions is considered to be significantly meaningful (Borg & Groenen, 2005).

While the findings do not support a disorganized semantic network for common animal names, they do suggest that AD participants have a reduced fund of words in the semantic network or weakened associations between less salient words. More specifically, while AD participants produced fewer category exemplars than the control group, a greater proportion of the exemplars were among the 11 high-frequency words selected for MDS analysis. Individuals with AD also made more repetitions on the category, but not letter, fluency task. Of these repetitions, more words were high-frequency exemplars. Since AD participants made significantly more repetitions only on the category fluency task, and because AD participants did not make more set-loss errors on either task compared with controls, the repetitions observed on the category task are thought to be more related to a decreased fund of words available in the semantic network in contrast to a perseverative process.

In support of this interpretation, repetitions made by AD participants have been found to correlate with deficits on tests of semantic knowledge (Lamar, et al., 1997). Ober and colleagues (1986) also found that a disruption in semantic memory processes did not affect AD participant’s ability to retrieve words with strong category associations even though fewer words in general were recalled. AD participants have also been found to produce a smaller number of low frequency or atypical words (Martin & Fedio, 1983; Sailor, Antoine, Diaz, Kuslansky, & Kluger, 2004; Tröster, Salmon, McClough, & Butters, 1989) and fewer novel semantic subcategories (Haugrud, Crossley, & Vrbancic, 2011) than controls. Therefore, a reduced semantic network may also explain why AD participants had significantly smaller cluster sizes on the category, but not letter, fluency task. Our findings suggest that deterioration of the semantic network integrity can occur separate from a disruption in the organization of the remaining accessible semantic words. As words become less salient or infrequently used, however, a breakdown in organization may become more prevalent. Given that individuals with AD have intact organization and access to high-frequency words it is suggested that common words may be better comprehended by these individuals and should be utilized when engineering interventions and in clinical and caregiving interactions.

One limitation to the current research is an inability to decipher whether it is more challenging to access less common semantic words due to attenuated associated connections, whether words in the semantic network are physically lost due to AD pathology, or whether it is a combination of the two. Chertkow and Bub (1990) found that individuals with AD rarely generate words on category fluency tasks for which they were unable to answer semantic-probe questions. This observation supports the notion of a degradation of the semantic memory store integrity. Future studies should continue to explore this question. The homogeneity of participants in regard to level of education and ethnicity is also a constraint to our study. Specifically, our study sample was composed largely of well-educated, Caucasian individuals, which limits our ability to generalize the results and conclusions to other populations with AD. Finally, the data used to complete this study was derived from two separate studies collected by the same laboratory at Washington State University. It is possible that differences in study administration and testing battery could have impacted the present results, although no systematic errors were evident.

Results from the present study revealed that individuals with AD are impaired on letter and category fluency in terms of total words and switches produced, with more notable deficits on the category fluency task. Clustering on category fluency was also reduced suggesting a deficit in the semantic network in addition to frontal lobe impairments. The Procrustes transformation made on the MDS plots revealed that the organizational strategies utilized by the mild-to-moderate AD group were comparable with healthy older adults for commonly produced animal exemplars. Given that the MDS provides a qualitative map that is open to interpretation, and different coordinates with equally good fit could have been derived based on a given starting configuration, future studies should subject MDS coordinates to a Procrustes transformation when making comparisons. The findings of similar response patterns across time along with the Procrustes transformation data suggest that the individuals with AD approached the fluency tasks in a similar manner as controls and exhibited intact organization of their semantic networks for high-frequency animal exemplars despite a reduced fund of words available. Overall, the data suggest that executive abilities involving search and retrieval processes and a reduced availability of semantically related words is contributing to the AD group’s poorer verbal fluency performance.

Funding

This work was supported by grants from the Life Science Discovery Fund of Washington State, National Institute of Bio Medical Imaging and Bioengineering [Grant #R01 EB009675], and Integrative Graduate Education Research Traineeship [Grant #DGE-0900781 2009-2014].

Conflict of Interest

None declared.
Acknowledgements

We thank members of the Aging and Dementia Laboratory at Washington State University for their help in collecting and scoring the data and Andy Weakley for his assistance with analyses involving MatLab.

References


