ELSEVIER

Neuropsychologia 44 (2006) 21–35

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Information content versus relational knowledge: Semantic deficits in patients with Alzheimer's disease[☆]

Justin M. Aronoff[a,*], Laura M. Gonnerman[b], Amit Almor[c],
Sudha Arunachalam[d], Daniel Kempler[e], Elaine S. Andersen[a,f,g]

[a] Hedco Neurosciences Building, 28c, University of Southern California, Neuroscience Program, Los Angeles, CA 90089, USA
[b] Department of Psychology, Lehigh University, Bethlehem, PA 18015, USA
[c] Department of Psychology, University of South Carolina, Columbia, SC 29208, USA
[d] Department of Linguistics, University of Pennsylvania, Philadelphia, PA 19104, USA
[e] Communication Sciences and Disorders, Emerson College, Boston, MA 02116-4624, USA
[f] Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA
[g] Department of Linguistics, University of Southern California, Los Angeles, CA 90089, USA

Received 15 April 2004; received in revised form 17 April 2005; accepted 25 April 2005
Available online 31 May 2005

Abstract

Studies of semantic impairment in Alzheimer's disease (AD) have yielded conflicting results, some finding evidence of considerable deficits, others finding that semantic knowledge is relatively intact. How do we reconcile findings from picture naming tasks that seem to indicate semantic impairment in AD with results from certain sorting tasks that suggest intact semantics? To investigate the basis of the contradictory results described above, we conducted a study using two types of tasks: (1) picture naming; and (2) board sorting. The board sorting task we used is a simultaneous similarity judgment task, in which participants are asked to place more similar concepts closer together and less similar ones farther apart. We compared the performance of AD patients on these two tasks, using a number of different analyses that yield very different patterns of results. Our results indicate that whether patients show impairment or not depends on both the nature of the task and the subsequent analysis chosen. Specifically, tasks and analyses that focus on relational knowledge (e.g., *dog* is more related to *cat* than to *camel*) lead to different conclusions than those based on specific information about individual items. These findings suggest that the board sorting method, when coupled with multiple analyses, provides a more complete picture of the underlying semantic deficit in AD than previous studies have shown.

Keywords: Semantic deficits; Board sorting; Similarity judgments; Language impairment

1. Introduction

Many studies have investigated the nature of semantic knowledge in patients with Alzheimer's disease (AD), but the results have often been mixed or even contradictory (e.g., Giustolisi, Bortolomeo, Daniele, Marra, & Gainotti, 1993; Gonnerman, Andersen, Devlin, Kempler, & Seidenberg, 1997; see Nebes, 1989 for a review). While some studies have found that patients with mild or even moderate AD exhibit semantic knowledge similar to that of normal controls (e.g., Bonilla & Johnson, 1995; Ober & Shenaut, 1999), others have found widespread semantic deficits (e.g., Nebes, 1989), and still others have observed selective semantic impairments that are restricted to certain tasks or domains (e.g., Chan, Salmon, & De La Pena, 2001). For example, some patients demonstrate difficulty in picture naming but not sorting tasks, or difficulties with only some semantic properties, such as domesticity (i.e., whether an animal is wild or domesticated), but not others, such as size (e.g., Chan, Butters, Salmon, & McGuire, 1993; Sacchett & Humphreys, 1992).

[☆] This study was carried out at the University of Southern California in the Departments of Psychology and Linguistics and the Program in Neuroscience

[*] Corresponding author. Tel.: +1 213 740 9221; fax: +1 213 740 5687.
E-mail address: aronoff@usc.edu (J.M. Aronoff).

0028-3932/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2005.04.014

Several factors may be responsible for the different results reported across studies. First, since there is typically high variability in AD patients' performance, the conflicting results may simply be due to different patient samples. Second, some previous studies, especially those that failed to find evidence for semantic deficits in AD patients, relied on restricted sets of items, often drawn from a small number of semantic categories (Bonilla & Johnson, 1995; Chan et al., 1993). In contrast, other studies, which have generally found evidence for semantic deficits in AD patients, employed larger sets of items drawn from a wider range of categories (e.g., Gonnerman et al., 1997). Because semantic impairments in AD may not be reflected in all semantic categories to the same extent (Gonnerman et al., 1997), studies that use a small set of items from a restricted range of semantic categories may miss deficits that are detected in studies that use a larger set of items drawn from a wider set of semantic categories. Finally, studies that failed to find evidence for semantic deficits typically used tasks that aimed to assess the overall organization of semantic categories in memory (e.g., Ober & Shenaut, 1999). An example is a board sorting task in which participants are asked to arrange a set of items from a semantic category according to their subjective similarity. In contrast, studies that found evidence for semantic deficits in AD tended to use tasks such as picture naming that assess performance on individual items (e.g., Barbarotto, Capitani, & Laiacona, 2001). Divergent results across these two types of studies could thus be the consequence of the different aspects of semantic processing that are assessed by each task.

The purpose of this paper is to explore the roles of item set, task differences, and type of analysis in determining whether semantic deficits are found, as well as the nature and basis of any such deficits. Following a detailed review of the tasks and analyses used in previous studies, we report the results of two experiments in which the same group of patients was tested on one large set of items drawn from a wide variety of semantic categories using two tasks that had previously yielded conflicting findings about AD patients' semantic deficits: picture naming and board sorting. Testing the same group of patients with the same items on these two tasks provides a powerful method for ascertaining how task and analysis differences can result in the seemingly conflicting findings found previously in the literature, while avoiding subject, item-, and category-specific biases. Using different tasks and analyses, we will show how apparently conflicting results can arise from the same group of patients, thus providing evidence that the various tasks and analyses reveal different aspects of the same semantic system, rather than patient population differences.

1.1. Picture naming

A number of researchers have used a basic picture naming task to investigate the semantic ability of patients with AD (Barbarotto et al., 2001; Chertkow & Bub, 1990; Laiacona, Barbarotto, & Capitani, 1998; see Nebes, 1989, for an overview). The general findings are that AD patients demonstrate a deficit when asked to provide the name for pictured objects or events. There is an active debate regarding whether these deficits represent damage to the semantic system or to some other system or aspect of lexical processing (Astell & Harley, 1998; Barbarotto, Capitani, Jori, Laiacona, & Molinari, 1998; Bayles, Tomoeda, & Cruz, 1999; Nakamura, Nakanishi, Hamanaka, Nakaaki, & Yoshida, 2000; Paganelli, Vigliocco, Vinson, Siri, & Cappa, 2003). For instance, some researchers have argued that naming deficits arise from lexical access difficulties (e.g., Astell & Harley, 1998, 2002; Somners & Pierce, 1990). Astell and Harley (1998) found that AD patients had impaired naming ability but relatively intact comprehension ability; they therefore argued that underlying semantic knowledge is still present but that access to the word forms themselves is impaired. In later work, Astell and Harley (2002) asked patients and normal controls to define words as a more direct test of semantic knowledge. The patients gave correct but impoverished definitions for many of the words. In addition, the patients did not show a preferential deficit for low frequency, atypical words (difficulty with low frequency words was a criteria for semantic deficits put forth by Warrington and Shallice (1979), and in fact gave more complete definitions for such items. Based on these findings, Astell and Harley argued that the deficit was a metalinguistic deficit, not a semantic one.

Other researchers have argued that the naming deficits arise from difficulties in the visual system. These researchers noticed a pattern of errors that led them to suggest that naming deficits are actually the result of a perceptual deficit, whereby difficulties in visually discriminating objects caused the misnaming (e.g., Kirshner, Webb, & Kelly, 1984).

However, some studies have shown that patients who clearly do not have perceptual deficits nonetheless have naming impairments (Chertkow & Bub, 1990; Huff, Corkin, & Growden, 1986). Additionally, although perceptual deficits may increase as the disease progresses (e.g., Hodges, Salmon, & Butters, 1991), semantically related errors continue to far outpace perceptual errors (Barbarotto et al., 1998). The systematic error pattern progression with increased levels of dementia also argues for a semantic explanation of the naming deficits. As the disease progresses, the error types shift from contrast coordinates to superordinates and non-responses, including responses such as "don't know" (Barbarotto et al., 1998; Paganelli et al., 2003; see Gonnerman, Aronoff, Almor, Kempler, & Andersen, 2004). There is no obvious lexical access explanation whereby errors would follow such a hierarchical pattern, whereas this pattern is easily explained by a progressive loss of semantic features.

Although lexical access deficits and perceptual difficulties likely contribute to naming deficits, there is a significant amount of data that is consistent with semantic deficits being primarily responsible for naming problems. First of all, as mentioned above, a large proportion of the errors that are made are semantically related to the target word (e.g., Barbarotto et al., 1998; Hodges et al., 1991). Second, relative proficiency in comprehension (as shown through picture pointing) compared to deficits in picture naming (Astell & Harley, 1998) is consistent with a damaged underlying semantic system since picture pointing, even when using within-category distracters, requires significantly less semantic information than picture naming. Third, in terms of the data presented by Astell and Harley (2002) which indicated that AD patients produce good definitions for low frequency, atypical items, they failed to take into account that typicality may affect preservation of features since atypical items often have very salient distinctive features (e.g., "a penguin does not fly" compared to "a dog barks"). Given that good definitions generally include distinctive features of the target item (Snow, 1990), it is not surprising that atypical items, with their salient distinctive features, would elicit good definitions even after substantial damage to the semantic system.

1.2. Similarity judgments

Similarity judgments have been used to investigate the organization of semantic knowledge in both healthy normal participants and AD patients (e.g., rating the similarity of *dog* and *cat* on a scale from one to seven). With normal participants, a large number of similarity judgments are usually obtained from a large group of participants and used to generate a representation of the participants' semantic space, assumed to represent the knowledge that the individuals have about the various concepts involved. While this method is effective with normal participants, it is often difficult to obtain large numbers of reliable similarity judgments from impaired populations such as AD patients.

Despite this difficulty, some researchers working with AD patients have elicited pair-wise similarity judgments, or used a triadic comparison task, a close variant in which participants must decide if one of three items is more similar to a second or a third item (e.g., is *dog* more similar to *cat* or *camel*). We will refer to both pair-wise and triadic comparison tasks as *serial* similarity judgment tasks here to distinguish them from board sorting, described in detail below. Using the serial similarity judgment task, Chan and colleagues (Chan, Butters, & Salmon, 1997) created multidimensional scaling (MDS) representations, based on responses from AD patients and normal elderly controls. Results indicated that AD patients were less consistent in their use of features (i.e., predation, domesticity, and size) than a comparable group of normal controls (NC). Also, unlike the NC group, the AD group tended to base similarity judgments on concrete (e.g., size) rather than abstract (e.g., domesticity) aspects of the meaning of test items. However, some studies using these types of similarity tasks have found normal performance (at least for certain categories) for AD patients. For instance, Chan et al. (2001) found that although patients demonstrated impairment for the category "animals," no impairment was found for the category "tools." Additionally, Rich, Park, Dopkins and Brandt (2002) found that AD patients performed similarly to normal controls on a sorting task where the number of piles used to sort the items was predefined. These apparent exceptions to the findings that patients with AD demonstrate semantic deficits in serial similarity tasks will be addressed in Section 4.

A major problem with the use of serial similarity judgments is that a very large number of trials is needed to obtain a complete set of similarity judgments for a small number of items, prompting some researchers to develop other methods to investigate within-category similarity. One such method is the board sorting task[1] (Bonilla & Johnson, 1995). In this task, multiple simultaneous similarity judgments are obtained by having participants place chips with words printed on them on a two dimensional grid, in a manner that represents the semantic relations between the concepts represented by those words, with words representing similar concepts being placed closer together than words representing less similar ones. The overall configuration of words on the board is taken to represent participants' semantic knowledge. Although this task is relatively complex in that it requires the simultaneous consideration of multiple similarity relations, it allows a relatively quick test of a large set of items from multiple semantic categories. In the following discussion we will refer to this kind of task as a *simultaneous* similarity task.

Bonilla and Johnson (1995) used this task to investigate the semantic deficits of patients with AD. Both NC and AD participants completed boards for two categories, one consisting of "animals" and a second consisting of "occupations". After completing each board, participants were asked to explain why they placed the pieces where they did. Based on the configuration of items on the boards (and aided by the descriptions of the participants), Bonilla and Johnson created an incidence matrix consisting of all possible pairs and a measure of whether or not the two members of a pair were placed in the same group by the subject. The incidence matrices for all participants in a given group were combined and an MDS analysis was conducted based on those data. The result was four composite boards for "animals" and four for "occupations" (one for each of four groups: normal controls, patients with mild AD, patients with moderate AD, and the combined group of patients with AD). They then visually compared the results from the normal controls and the patients with AD. In contrast to the findings of Chan and colleagues from serial similarity judgment tasks, Bonilla and Johnson found that their patients with mild AD (and to a lesser degree their patients with moderate AD) grouped the words in a manner similar to that of the normal controls. They used these findings to argue that semantic knowledge is largely intact in AD patients.

In a similar board sorting task, Ober and Shenaut (1999) asked participants to place words written on tags attached to wooden pegs in a pegboard, again with more similar items

---

[1] The term board sorting rather than simply sorting is used here to distinguish this class of experiments from those where participants simply group stimuli.

being placed more closely together. They found that the representations of patients with Alzheimer's disease were generally similar to those of normal controls. From these results they argued, as did Bonilla and Johnson, that the semantic knowledge of patients with AD is largely intact.

How do we reconcile the findings from picture naming tasks, which seem to indicate that patients with AD have impaired semantic representations, with the results of the board sorting tasks, which seem to indicate that they do not? Even more puzzling is why the results from different types of similarity tasks do not agree, with serial similarity tasks generally showing deficits similar to picture naming, but simultaneous similarity tasks suggesting intact semantic representations. To investigate the basis of this contradiction, we conducted a study using both picture naming and board sorting tasks. In what follows we present a comparison of the performance of AD patients on these two tasks, providing a number of different analyses that yield very different results; under certain analyses, patients' semantic knowledge looks impaired, while other analyses suggest that their knowledge is relatively intact on the same task. More interestingly, we can predict whether the patients show impairment or not based on whether the task and subsequent analysis examine relational (e.g., *dog* is more related to *cat* than to *camel*) or item-specific (e.g., *dogs bark*) information. In Section 4 we will explain how this pattern of behavior can be accounted for with the type of distributed semantic-feature model we have previously described (Devlin, Gonnerman, Andersen, & Seidenberg, 1998; Gonnerman et al., 1997).

2. Method

2.1. Participants

The participants were 64 individuals who were paid for their participation. The group of patients with AD included 15 individuals who were diagnosed with probable Alzheimer's disease using the NINCDS-ADRDA criteria (McKhann et al., 1984). Results of neurological, laboratory (including computer tomography or magnetic resonance scan), and neuropsychological assessment failed to suggest other causes of dementia. The old normal (ON) group consisted of 24 elderly individuals (mean age 78). The young normal (YN) group consisted of 25 undergraduate students from the University of Southern California. For most analyses the two NC groups were combined and referred to as the YN/ON group. All participants were native speakers of standard American English. See Table 1 for age, years of education, and Mini-Mental State Exam (MMSE) score information. All recruitment and testing procedures were approved by the University of Southern California's Institutional Review Board (IRB). All participants signed an informed consent form approved by the IRB. For the patients with Alzheimer's disease these forms were signed by a legal guardian.

Table 1
Characteristics of the participants in each group

Group	*N*	Mean age (S.D.)	Mean education (S.D.)	Mean MMSE (S.D.)
YN	25	20.2 (2.5)	14.3 (1.4)	29 (0.6)
ON	24	78.1 (5.3)	16.8 (2.5)	29 (1.1)
AD	15	83.5 (3.8)[*,†]	14.4 (2.9)[*]	20 (3.3)[*,†]

[*] Significantly different than ON ($p < .05$).
[†] Significantly different than YN ($p < .05$).

2.2. Procedure

Participants completed the picture naming and board sorting tasks as part of a larger study. One hundred and 44 nouns were used across the various experiments, representing a diverse set of concepts from 12 semantic categories. The items were selected from both natural kind and artifact categories and controlled for familiarity, imageability, frequency (Francis & Kučera, 1982; Snodgrass & Vanderwart, 1980; Wilson, 1988), and typicality across domains (Battig & Montague, 1969). The 12 sets of concepts (each consisting of 12 words) are presented in Appendix A.

For the picture naming task, participants named a series of color photographs presented one at a time on a computer screen. All participants were asked to name all 144 pictures. They were given as much time as necessary to answer, and their answers were recorded for later transcription.

The board sorting task was adapted from the one used by Bonilla and Johnson (1995). Participants were presented with sets of 12 printed words on 1″ × 1″ foam-board chips. They were instructed that the goal of the task was to place similar words close together and less similar words farther apart. To supplement the instructions, an example board using colored chips was completed by the experimenter. As the example board was being completed, the experimenter explained why she placed the chips where she did (e.g., "I placed red next to orange because I think they are very similar. I placed blue farther away because I think it is less similar to red than orange"). All participants were given the same example using the same chips and distances. The participant was then given a set of either colored chips or chips with words printed on them and instructed to study them before placing them on the board. Participants were asked if they understood the instructions and, if needed, the instructions were repeated. The task was completed on a laminated board with a 10 × 10 square grid.

Each participant completed four boards, selected from the natural kinds and artifacts categories (see Appendix A for the number of participants per group that completed each category). For the AD group, the categories were determined based on the participant's ability to name the concepts in the picture naming task. Two of the categories chosen were ones that the participant had relative difficulty naming during the picture naming task (poorly-named boards) and the other two were ones the participant was relatively more proficient at naming during the picture naming task (well-named boards).

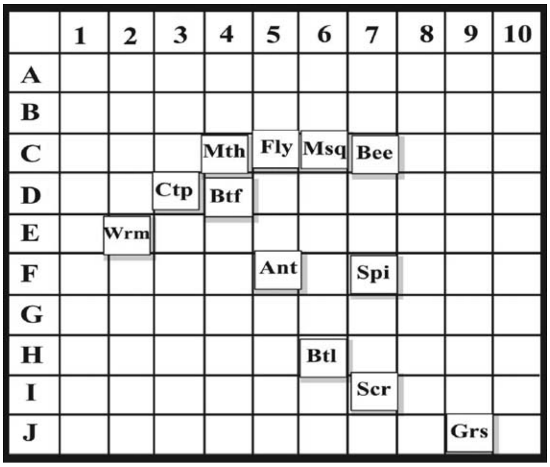

Fig. 1. An example board from the board sorting task for the category insects, including moth (Mth), fly (Fly), mosquito (Msq), bee (Bee), caterpillar (Ctp), butterfly (Btf), worm (Wrm), ant (Ant), spider (Spi), beetle (Btl), scorpion (Scr), and grasshopper (Grs); all words were spelled out in the actual task and are abbreviated here for figure legibility. Note that more similar concepts are placed close to one another and less similar concepts are placed farther apart.

The mean error rate was 25% for the well-named boards and 61% for the poorly-named boards. Four boards were also assigned to each participant from the YN and ON groups. These were assigned pseudo-randomly, with each participant receiving boards from two natural kinds categories and two artifacts categories. In addition, all participants completed a board with a set of 12 colored chips (none of which were used in the example board) that did not have any words written on them. The color board functioned as a control to verify that the differences in AD performance were not caused by an inability to complete the task.[2] An example board is shown in Fig. 1.

3. Results

3.1. Picture naming

Naming responses were coded as correct, incorrect, or machine error. Synonyms that were provided by normal control participants (e.g., sofa for couch) were included as correct for all participants as they demonstrate intact semantic knowledge of the intended concept. Two YN participants did not complete this task.

Results revealed that YN controls correctly named 86% of the pictures, ON controls 85%, and AD patients 62%. *t*-tests were conducted comparing all three groups. To control for family-wise error, alpha was adjusted using Rom's method (Rom, 1990). The results demonstrated a significant difference in picture naming scores between YN and AD (t (15) = −4.15, $p < .0009$), and between ON and AD (t (16) = 3.95, $p < .002$), but no significant difference between YN and ON (t (44) = −.48, n.s.).

3.2. Board sorting

To analyze the sorting data, each board was first converted into a set of 66 data points (comparing each of the 12 items to every other item yields 66 data points) representing each of the Euclidian distances between any two chips on the board. The unit of measurement used was board spaces: thus, the closest together any two chips could be placed on the board was one space apart; the farthest two chips could be from each other on the board was 12.73 spaces.[3] It is important to note that the degree of similarity between the items in one category is not necessarily the same as the degree of similarity between the items in another category. Although this does not pose a problem when analyzing the data one category at a time, certain analyses required looking at the data from all the categories combined. Because of the way categories were selected for AD participants, it was not possible to match the normal controls with the AD group in terms of the number of participants that completed each board. Doing so would have prevented us from maintaining a representative distribution of boards for any analyses that only involved the normal controls. Therefore, when comparing the complete data set between groups, an average was created per board per group regardless of the number of participants in each group that completed a given board. Thus each category received equal weighting. Because of the selection criteria for the AD group, one category (local wildlife) was not completed by anyone in this group, and thus this category was eliminated from all analyses involving the board sorting data. Additionally, one board from one YN participant could not be analyzed because of an experimenter error.

3.3. Color boards

To verify that the patients with AD were capable of performing the board sorting task at a level comparable to YNs and ONs, such that any difference in results between the groups in sorting semantic categories could be attributed to semantic deficits, we compared the performance of the AD group with that of the YN/ON group for a version of the

---

[2] Although a number of studies have suggested that semantic deficits can affect categorization of colors (e.g. Davidoff & Roberson, 2004; Roberson, Davidoff, & Braisby, 1999), similarity judgments appear to remain intact (Roberson, Davidoff, & Braisby, 1999). Given that our task does not require explicit categorization, but rather depends on similarity relations, we expected it to resemble similarity judgment tasks in that AD patients would not show an impairment with color sorting. This expectation was indeed confirmed by our results as described below.

[3] This is derived using the Pythagorean theorem. The distance from a chip in the bottom right corner to one on the top right corner is 9, and the distance from a chip in the bottom right corner to one in the bottom left corner is also 9, thus the distance from a chip in the bottom right corner to one in the top left corner is the square root of 9^2 plus 9^2, or 12.73.

board sorting task that relied significantly less on semantic information, using colored chips instead of the word chips. The distance for each pair of colors was averaged across all the participants in each group. Thus, for each group there were 66 data points (one per item pair). Data from two YN participants and one AD patient could not be included due to experimenter error. The correlation between the AD and YN/ON group data was calculated and a highly significant correlation was found between the two sets of data ($r = .82$, $p < .001$). This suggests a high similarity between the color board for the AD group and that for the YN/ON group. Additionally, to further verify that the two groups' color boards were comparable, we conducted a paired *t*-test (again with 66 data points per group) and found that there was no significant difference between the two ($t\ (65) = -.95$, n.s.). These combined results confirm that the AD patients were able to complete the board sorting task in a manner similar to the YNs and ONs.

3.4. Relational knowledge

A number of previous studies have used analyses that look at the relational properties among the various items (i.e., is chip A closer to chip B or chip C?), instead of ones that compare the exact similarity of two items across groups (i.e., how close is one chip to another?). We will refer to measures looking at exact similarity as measures of information content as similarity has been suggested to reflect the degree to which the semantic information in two concepts overlaps (e.g., Gentner, 1983; Tversky, 1977). This difference is important in that two boards could be identical in terms of one measure and very different in terms of the other. For instance, if one participant rated the similarity of *dog* and *cat* as 1 and the similarity of *cat* and *pig* as 3 and a second participant rated the similarity of *dog* and *cat* as 2 and the similarity of *cat* and *pig* as 6, a measure of relational properties (such as a correlation) would consider the two participants to have identical representations (i.e., strongly correlated), whereas an analysis measuring the exact similarity of the items (such as mean distance) would find that the two participants had radically different semantic representations. To examine the degree of both relational knowledge and information content in each participant's semantic system, two classes of analyses were chosen. The first class of analyses, discussed in this section, examined differences in relational knowledge across the groups. The second class of analyses, discussed in a later section, examined differences in the information content for concepts in the various categories across the groups. Two methods of analysis of relational properties were selected, one measuring the overlap between groups of the most closely related pairs, using the Pathfinder algorithm, and the other examining the correlation between the chip placement for the YN/ON group and ADs. Both of these analyses use methods that rely on relative placement of items and thus are well equipped to examine the relational knowledge that the AD group possesses. Each of these analyses is described in more detail below.

3.4.1. Pathfinder

To compare the results from the present data with those from the Ober and Shenaut (1999) study, a pathfinder analysis (Bonilla & Johnson, 1995; Chan et al., 1997; Ober & Shenaut, 1999; Schvaneveldt, 1990) was conducted. The Pathfinder algorithm is a method that can be used to select the most relevant pairs in a set of interrelated concepts. It works by selecting the pairs with the shortest distance (i.e., greatest similarity) between them (either directly or through a path) while maintaining complete interconnectivity among all the concepts. That is to say, it minimizes the mean distance for the set of items while maintaining a path whereby all items are connected to all other items. See Fig. 2 for an example of the Pathfinder analysis.

Following Ober and Shenaut (1999), Pathfinder networks were created for each group (YN, ON, and AD) and category (with $r = \infty$ and $q = N - 1$). The parameters r and q were set such that the length of a path was simply the length of the link with the greatest distance in the path, a necessary requirement when using distances drawn from two-dimensional representations. Additionally, a path could pass through every item if necessary. This allowed for the sparsest possible network while maintaining paths between all items. Using this analysis (and these parameters), Ober and Shenaut found that there was a significant amount of overlap between the pairs that the algorithm selected based on the NC's data and those based on the AD patients' data for the category "animals," but not for the category "instruments." It should, however, be noted that a separate group of healthy normal participants did rate the pairs selected based on the AD data for "instruments" to be as reasonable as those based on the NC data. Ober and Shenaut interpreted these results as indicating that the semantic knowledge of the AD group was similar to that of the NC group. Importantly, since the Pathfinder algorithm uses ranked rather than absolute distance, these results should be interpreted as indicating a general similarity between the relative placements of the various concepts used in the task. For example, this analysis shows that *dog* is more similar to *cat* than to *camel* for both groups. However, this analysis cannot be used to determine whether or not the amount of information about each concept is the same for both groups, since it does not distinguish between identical boards and boards where the general configuration of chips is the same but the actual number of spaces between them is not.

To verify that the patients in our study demonstrated a similar pattern of results as those in the study by Ober and Shenaut (1999), each composite AD pathfinder network (one per category) was compared with the corresponding composite YN and ON network. A composite YN/ON network was generated for later analysis. To interpret the pathfinder networks we used a method similar to that used by Ober and Shenaut (1999). In so doing, the number of shared over total links[4] (intersection:union ratio) was calculated between

---

[4] If the pathfinder network for group A had the links *dog*:*cat*, *dog*:*pig*, and *horse*:*camel* and the pathfinder network for group B had the links *dog*:*cat*,

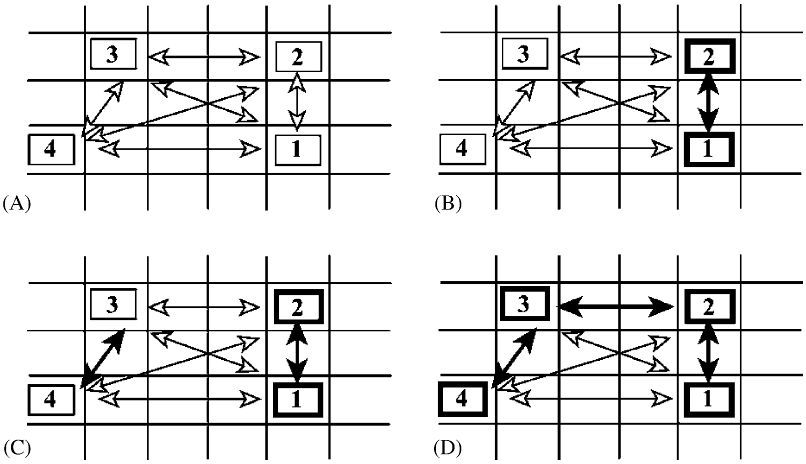

Fig. 2. This is a schematic representation of the pathfinder algorithm. The goal is to create a path with the minimal length of combined links from one item to all other items (in this case, from item 1 to item 4 in sequence). (A) A simplified example board with four items is shown with all the possible links (outlined arrows). (B) To conduct the pathfinder algorithm, the shortest links are added to the network (signified by dark arrows) and it is determined whether you can pass from item 1 to all other items using only the added lines As each item is reached along the path, it receives a dark outline. With the added link, only item 2 is reachable from item 1, and thus more links need to be added. (C) If all items cannot be reached using the available links, the next shortest links are added and it is determined if a complete path exists. In this case, since the only path available from item 1 stops at item 2, no additional items are reached and more links are needed. (D) Links continue to be added in the order of their length (shortest to longest) until a complete path exists, linking one item to all other items.

pairs of participants on a particular board. This overlap ratio was then normalized using an arcsine transformation. To calculate chance overlap, a set of 501 random boards was generated by randomly placing the same 12 chips on each board. Pathfinder networks were then created for each of these random boards and the percentage of overlap between the Pathfinder networks of each set of adjacent boards (number of pairs in both divided by the number of pairs in either) was calculated. This method resulted in 500 overlap scores representing chance performance. These scores were used to create a .95 confidence interval for chance overlap consisting of the 13th lowest and highest overlap scores (i.e., $p < .05$). The overlap scores were then also normalized with an arcsine transformation resulting in a .95 confidence interval of (.045, .15). The normalized mean overlap ratio across all categories was calculated separately for AD with YN and ON. Both mean overlap ratios were greater than the confidence interval derived for chance (.26 and .20, respectively), indicating a significant amount of agreement across the groups. Our results using this method are thus similar to those found by Ober and Shenaut (1999).

To investigate performance on individual categories, pathfinder networks were created for each participant and each board. Overlap ratios were calculated for each AD patient with each NC, as well as for each YN with each ON. As Ober and Shenaut (1999) found, approximately half of the categories showed a significant difference between AD versus YN/ON overlap and YN versus ON overlap. See Table 2. This suggests that, despite the finding that the data from the AD and YN/ON groups were more similar than chance, the AD group was not as similar to the YN/ON group as the two NC groups were to each other.

To determine how well the data from each AD patient would fit into the composite YN/ON network, the overlap ratio between each AD patient's network and the composite YN/ON network was calculated for each category. These ratios were then averaged within a category and compared with the chance scores described above. This comparison revealed a significant similarity between the AD networks and the composite YN/ON board for nine out of 11 categories (all but "zoo animals" and "clothing"). See Table 2. These combined results suggest, as did the findings of Ober and Shenaut (1999), that the semantic knowledge demonstrated by AD patients on the board sorting task was significantly similar to

Table 2
Similarities and differences in pathfinder networks across groups for each category

Category
Domestic animals†
Insects*,†
Kitchen items†
Musical instruments*,†
Carpenter's tools*,†
Toys†
Vehicles†
Fruits*,†
Vegetables†
Clothing*
Zoo animals

* Significant difference between YN/ON × AD and YN × ON agreement ratios ($p < .05$).

† Significantly more overlap between YN/ON and individual AD results than that obtained with random boards ($p < .05$).

---

and *duck*:*goose*, then the ratio would be 1/4 since one link occurs in both networks (*dog*:*cat*) and there are a total of four unduplicated links between the two networks (*dog*:*cat*, *dog*:*pig*, *horse*:*camel*, and *duck*:*goose*).

that of the normal controls. However, the performance can not be considered normal, since AD patients' networks often differed from the YN/ON networks more than the YN networks differed from the ON networks.

3.4.2. Correlation

In addition to the Pathfinder analysis, we also used a second metric of agreement in terms of relative semantic similarity across groups, based on the correlations between the pair-wise distances in the YN/ON and AD groups. This method measures the relative similarity of two concepts, regardless of the exact distance between them on the board. Much like the pathfinder-based measures, this measure is sensitive to relative distances such as when *dog* is placed more closely to *cat* than to *duck* for both the YN/ON group and the AD group, even if one group says that *dog* is one space away from *cat* and the other says they are two spaces apart. Mean distances were calculated for each pair for a given category based on the original 66 data points for each participant in a given group. Thus each category was represented by 66 data points per group, which represented the mean distance apart for each pair-wise comparison for that category and group.

Pearson's correlation was first conducted using the average distances from each group for all 726 pairs (66 pairs per board across 11 boards), with all three correlations significant (YN × ON: $r = .65$, $p < .001$; YN × AD: $r = .26$, $p < .001$; ON × AD: $r = .33$, $p < .001$), indicating that the relative configurations among concepts for the AD and YN/ON group were significantly more similar than chance. Even though these results indicated that the AD patients' concepts are arrayed in a way similar to normal controls, this measure did demonstrate a difference between normal controls and AD participants. Although a number of methods have been proposed for comparing correlations across independent groups (e.g., Duncan & Layard, 1973; Yu & Dunn, 1982), recent investigation of these methods has found that the most effective measure is Bootstrap comparisons[5] (Wilcox & Muska, 2002). Therefore, Bootstrap comparisons were conducted to determine whether the data from AD participants were more weakly correlated with normal controls than the data from YNs with ONs. The results indicated that YNs were more strongly correlated with ONs than with ADs, (.95 *confidence interval* is .14, .32) and ONs were more strongly correlated with YNs than with ADs (.95 *confidence interval* is .14, .33). However, no significant difference was found between the correlation of YNs with ADs and ONs with ADs (.95 *confidence interval* is −.08, .12).

Table 3
Categories for which there is a significant correlation for the placement of AD and YN/ON chips

Category	AD×	r
Domestic animals	$\text{YN}^{**,\dagger}$	.57
	$\text{ON}^{**,\dagger}$	.59
Zoo animals	YN	.20
	ON	.19
Insects	YN	.14
	ON	.24
Fruits	$\text{YN}^{*,\dagger}$	.29
	$\text{ON}^{**,\dagger}$	.41
Vegetables	YN	−.05
	ON	.13
Musical instruments	YN	.14
	$\text{ON}^{**,\dagger}$	.42
Carpenter's tools	$\text{YN}^{**,\dagger}$	.38
	$\text{ON}^{**,\dagger}$	.43
Vehicles	$\text{YN}^{**,\dagger}$	.49
	$\text{ON}^{**,\dagger}$	.44
Clothing	$\text{YN}^{**,\dagger}$	.40
	ON	.19
Kitchen items	$\text{YN}^{**,\dagger}$	.47
	ON^{*}	.25
Toys	$\text{YN}^{**,\dagger}$	.37
	$\text{ON}^{*,\dagger}$	.29

[*] $p < .05$.
[**] $p < .01$.
[†] Correlation between AD and NC boards greater than that obtained with random boards ($p < .05$).

An additional analysis was conducted to determine if the correlation also existed for boards representing single categories. The average distance for each pair of items for a given category was calculated for each participant group and category and served as the data for the analysis. As Table 3 indicates, there was a significant correlation between the data from the AD group and those from at least one of the normal control groups for eight out of the eleven categories.

To verify that both the group and category correlations were higher than chance given the dynamics of the board, random boards were generated. Because the original mean scores for each category were calculated based on the average score for each pair of concepts for approximately ten subjects, groups of ten boards were averaged. In order to create a representative set of 500 correlations based on adjacent pairs of averaged boards, 5010 random boards were generated. Correlations were then calculated between serial pairs of composite boards, resulting in 500 correlation coefficients based on a random distribution. This was then used to calculate a confidence interval consisting of 95% of the correlation coefficients (−.265, .269). Using this confidence

---

[5] In order to conduct this analysis, data points are randomly selected from each data set to create two sample data sets, one set taken from each distribution of the actual data. This is done a number of times, creating approximately 600 sample data sets from each distribution, which are presumably similar to those from which the original data was obtained. Correlations are then conducted using the various sample sets. The difference is obtained between correlation coefficients from each distribution and these are ordered by magnitude. A confidence interval is calculated based on the interval within which 95% of the data is contained. If this interval contains 0 (i.e., if less than 95% of the correlation pairs have a difference that is to a given side of 0), then it is determined that there is no significant difference between the correlations (see Wilcox, 2003).

interval, the correlation between ON and AD distances was significantly greater than chance given the restrictions of the task ($r = .83$). Additional analysis of the distribution indicated that the correlation between YN and AD was marginally significant ($r = .26$). All the boards that were found to have a significant correlation when compared to a normal group had a correlation coefficient above the chance confidence interval (with the exception of “kitchen items” for the ON group) indicating that, in general, the correlations, and by extension the semantic space, were significantly similar for the AD and YN/ON groups. In the following section another class of analyses will be used to further investigate if the semantic representations are truly similar, despite the similarity in relative relatedness of concepts across the two groups.

3.5. Absolute relations

As in previous studies using board sorting tasks, our results indicated that AD patients’ performance was similar to that of normal controls for a number of categories. However, these analyses only demonstrated that the relative placement of the concepts (e.g., *dog* is closer to *cat* than it is to *camel*) was similar across populations. It is possible that other classes of analyses may find greater differences between the AD group and the YN and ON groups. To examine this possibility we conducted an analysis aimed at investigating the amount of semantic information available to patients with AD. Instead of trying to determine if the patients knew simply that *dog* is more similar to *cat* than to *camel*, the analysis was aimed at determining if the patients knew exactly how similar *dog* and *cat* are to each other. The analysis consisted of measuring the mean distance between all the chips and comparing this across groups.

3.6. Spread

To explore differences in the absolute distance between various concepts on the board, we calculated a measure we will call *spread*. Spread consists of the average distance between pairs of words over an entire board. As demonstrated below, this measure reflects the amount of knowledge a participant has about concepts in a given category. The rationale behind this measure is as follows: each category can be divided into a series of sub-categories. For instance, *insects* can be grouped by *manner of motion* (with *flying* and *crawling* insects in separate groups), or they could be grouped by *number of legs* (e.g., *spiders*, *ants*, and *flies* versus *caterpillars* and *centipedes*). *Crawling insects*, in turn, can be divided into *crawling insects with four or fewer legs*, and *crawling insects with more than four legs*. *Crawling insects with more than four legs* can be divided into *those with a countable number of legs*, and *those with too many legs to count*. These meaningful subcategories allow for the formation of clusters and subclusters on the board (e.g., one cluster for *flying insects* and one for *crawling* ones, with subclusters in the *crawling* cluster based on *number of legs*). The greater the number of meaningful subcategories, the more separate clusters and subclusters will appear on the board. In general, as the number of individual clusters increases, the spread score also generally increases because non-contiguous clusters add blank spaces between groups of chips, thus adding to the spread score. Therefore the measure of spread on each board is one way to calculate and compare the degree of semantic elaboration within a conceptual category for a given participant.[6] Although it is possible to increase spread simply by distributing the same clusters over a larger space, indicating a different strategy rather than a difference in knowledge, results presented below suggest that that was not the case for our data.

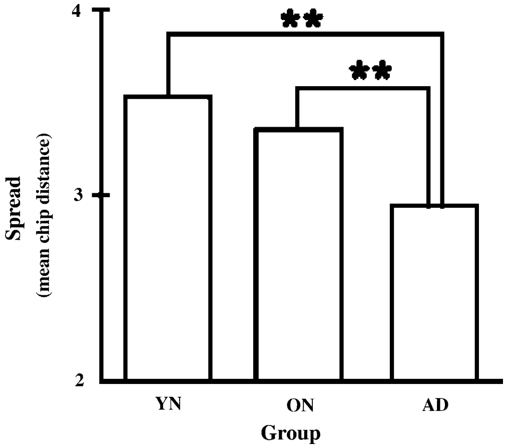

Fig. 3. Mean distances between chips for YN, ON, and AD groups on word boards (YN × AD: $t\,(20) = 4.33$, $p < .0003$; ON × AD: $t\,(14) = 4.03$, $p < .001$; YN × ON: $t\,(14) = 1.55$, $p > .1$).

To compare the performance of AD patients to YN and ON controls on the word chips, *t*-tests were used comparing each group, using Rom’s method (Rom, 1990) to adjust for multiple comparisons. Each participant contributed one data point consisting of the average spread score across all the boards they completed. Data from the AD group showed a significantly smaller spread than the YN and the ON groups (YN: $t\,(20) = 4.33$, $p < .0003$; ON: $t\,(14) = 4.03$, $p < .001$), with no significant difference in spread between the YN and ON groups ($t\,(14) = 1.55$, $p > .1$). See Fig. 3.

It is possible to interpret these results as indicating that AD patients just have general difficulty with the board sorting task, or that their strategy for such a task is to simply place all the chips close together. To determine if this was the case, we analyzed the data from the color board with the

---

[6] We elected to use this method rather than using a cluster analysis because of the exploratory nature of cluster analyses and our desire to quantitatively analyze the data. Unlike most kinds of cluster analyses, this method allows us to compare individual boards rather than relying on a composite board for a group, thus providing a way to analyze the semantic performance of individual participants. Additionally, because this method provides a quantitative measure that is independent of the specific semantic category of any given board, it also allows for data to be compared across categories.

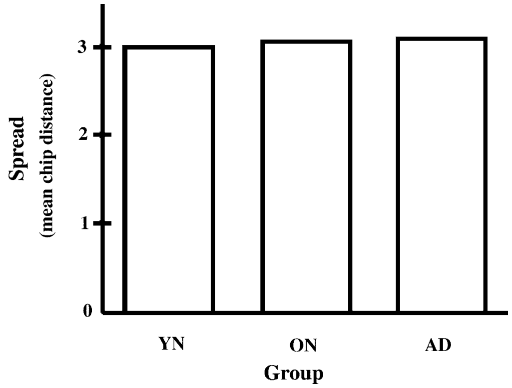

Fig. 4. Mean distances between chips on the Color board for YN, ON, and AD groups (all differences non-significant).

spread measure (recall that this task relies heavily on simple visual similarity rather than semantic knowledge). The data for each group consisted of the mean distance between all the chips on the color board for each subject. The results indicated that AD patients did not have a tendency to place chips closer together than the YN/ON group for color boards (YN mean: 3.02, ON mean: 3.11, AD mean: 3.13, no significant difference), see Fig. 4. We can therefore be confident that the AD patients did not have a general strategy of placing chips closer together than the YN/ON group. This suggests that any differences between AD patients and normal controls were due to underlying differences in semantic knowledge and not simply to task demands.

However, it is still possible that other non-semantic factors (e.g., overall difficulty or strategy differences that are only relevant with words) may influence performance on the word boards and still allow for normal behavior on the color board. In order to further verify that spread is related to conceptual knowledge, each AD participant's boards were divided equally and placed into one of two groups based on their picture naming score for that category, such that two categories for each patient were labeled as well-named and two as poorly-named. The spread score was then calculated for the well-named boards and the poorly-named boards, with each patient contributing two data points (i.e., the mean distance between all the chips for a given board) to each. Importantly, since each participant's set of boards was divided equally into each group during analysis, each participant served as his or her own control. Thus, any semantic impairments shown in the task were not confounded with other factors that differed across patients, such as disease progression and working memory deficits, since these factors were identical for both the well-named and the poorly-named sets of boards. Much like the difference between YN/ON and AD spread, there was a significant difference between the well-named boards and the poorly-named boards, such that categories that participants named more accurately also had a greater spread (t (29) = 3.46, $p < .002$), providing further evidence that spread is related to semantic knowledge, and not simply to differences in strategy. See Fig. 5.

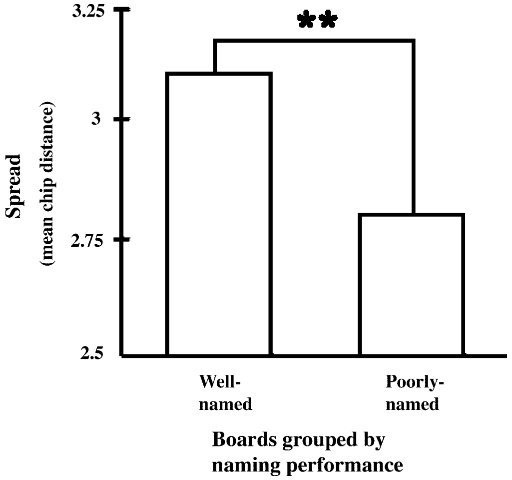

Fig. 5. Mean distances between chips for well-named and poorly-named boards for the AD group (t (29) = 3.46, $p < .002$).

3.7. Comparing relative and absolute measures

As the pathfinder results above demonstrate, pathfinder networks for the ON/YN and AD groups were significantly similar for only nine out of 11 categories, and their boards were significantly correlated for only eight out of 11 categories. This opens up the possibility that, while there was a difference in spread between YNs/ONs and ADs, this difference may have arisen solely from the categories that showed a difference between the AD and YN/ON groups in the pathfinder or correlation analyses. To test this we reexamined spread using only the categories that showed significantly similar YN/ON and AD pathfinder networks and, in a separate analysis, those that showed a significant correlation between the YN/ON and AD groups. In both of these restricted subgroups of categories, the AD group still showed significantly smaller spread than the normal controls (Pathfinder: YN/ON × AD: t (8) = 3.8, $p < .005$; Correlation: YN/ON × AD: t (7) = 3.4, $p < .012$).

4. Discussion

Consistent with the findings of Bonilla and Johnson (1995) and Ober and Shenaut (1999), the Pathfinder analysis of the board sorting behavior of our mild-moderate AD participants suggests that, for most categories, patients' knowledge of relative similarity among concepts was not significantly different from that of normal controls: when, as in previous studies, relational similarity was compared to chance, the majority of boards from the AD group were indistinguishable from those from the YN/ON groups. However, a closer inspection revealed that the relative similarity was less consistent between the AD and YN/ON groups than between the YN and ON groups. Moreover, further analysis of the boards revealed that, although the relational knowledge of the AD participants was similar to that of the normal controls, the AD and YN/ON groups differed dramatically in terms of spread, a measure of information content.

Our analyses demonstrate that even when the same group of participants is used, and when a large set of items from a representative set of semantic categories is used, task choice and analysis type can lead to different views about semantic impairments in AD. Although, as we have shown, some of the apparent contradictions can be attributed to whether the task and analysis aim to assess relational versus absolute semantic knowledge, we have not yet addressed the question of why AD patients' semantic performance would appear worse in absolute tasks and analyses than in relative ones. Answering this question requires a careful consideration of task demands and analyses from a theoretical perspective. Therefore we now present a theoretical framework that will allow us to address the methodological issues with reference to the nature of the organization of semantic memory.

4.1. Featural approaches to semantic representations

A number of researchers have proposed that concepts are, at least in part, composed of sets of interconnected features (e.g., McRae, de Sa, & Seidenberg, 1997; Shallice, 1988; Smith & Medin, 1981; Vinson, Vigliocco, Cappa, & Siri, 2003). In addition, others have suggested that some concepts are related through an overlap of features (e.g., Gentner, 1983; Tversky, 1977). Although a number of other theorists have recently questioned the validity of this approach (Hayes & Bissett, 1998; Williams, 1996; see Hutchison, 2003, and Lucas, 2000 for discussion), both behavioral and modeling studies of impaired populations have provided considerable support for the featural account (e.g., Devlin et al., 1998; Gonnerman et al., 1997; McRae et al., 1997; Vinson et al., 2003). For instance, Gonnerman et al. (1997) showed that artifact and natural kind categories demonstrate different patterns of deterioration, with artifact categories showing slight deficits compared to natural kinds early on, followed by more pronounced deficits in natural kind categories as the disease progresses. The authors explained this pattern of impairment by noting differences in both the number and type of intercorrelated features underlying natural kinds versus artifacts. Namely, natural kinds categories have relatively more perceptual features that are more highly intercorrelated (e.g., most *animals* that *have a beak* also *fly*). These intercorrelations help resist the effects of damage because features that become unavailable for one concept can be indirectly activated through related items (see also Aronoff, Gonnerman, Andersen, Kempler, & Almor, 2003). For example, if *has a tail* is no longer directly available for *dog*, it may be indirectly available through the knowledge that a *cat has a tail* and a *cat* shares many features with a *dog*. Thus the observed deficit pattern is consistent with an approach where semantic features are affected by random damage in AD.

The account of the pattern of deficits discussed in Gonnerman et al. (1997) was further supported by simulations from a connectionist model by Devlin et al. (1998). This model assumed the structure of semantic space proposed in Gonnerman et al. (1997) in which concepts are made up of a variety of features dispersed across the semantic space. Connections to individual features were randomly deleted across concepts and the pattern of semantic deficit from Gonnerman et al. (1997) was in fact found (see also McRae et al., 1997; Vinson et al., 2003).

Under the featural approach, there is a distinction between the amount of detail in a particular representation of a concept and how that concept relates to other ones.[7] Since the relationship between two concepts within a category is based on the amount of overlap between them compared with the amount of overlap between another pair of concepts, this structure is relative, rather than absolute. Thus, if a number of features become unavailable across various concepts, the amount of information within those concepts will become diminished, but the overall relation between all the various concepts may remain roughly intact. See Fig. 6 for an illustration.

4.2. Task demands

In comparing the results we obtained with those reported in the literature, it is important to consider differences in task demands because different tasks require different amounts and types of semantic knowledge (as well as other abilities), and thus might naturally lead to differing profiles of ability or impairment. For instance, a picture pointing task, where participants are asked to point to the named picture among a number of foils, may be completed successfully even in the face of considerable semantic impairment (depending on the foils used) because a participant need only have sufficient semantic knowledge to distinguish the target from the foils. Theoretically, in a case where the target is *zebra* and the foils are *horse*, *candy cane*, and *giraffe*, a participant who only retained the semantic features *has stripes* and *has legs* might be able to successfully choose the target without knowing anything else about *zebras*.

In contrast to picture pointing, picture naming requires more semantic information, though again, completely intact semantic knowledge is not required. For instance, if only the feature *has stripes* and *has legs* are available in a picture naming task for *zebra*, there are a number of objects to which these features can apply (even if they are not features which that concept must necessarily have, as is the case for *cat*, which can *have stripes* even though not all cats do). However, there are many features, such as *has eyes*, that could be damaged, but this deficit would not affect the naming of the object, and thus the accuracy rate in naming might be the same as that for someone with that knowledge intact.

---

[7] This is not to say that there is a complete division between the two; for instance, a vacuous concept (i.e., one with no features) has no relation to any other concept. This division only applies to representations of concepts with at least a minimum number of features, although the exact number is difficult to determine.

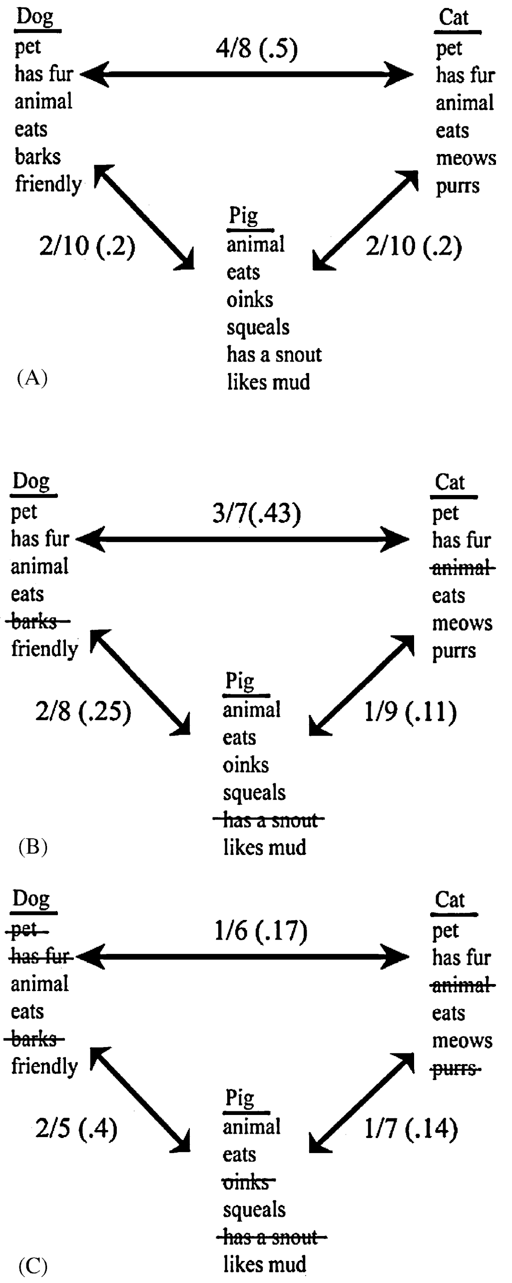

Fig. 6. Example of a potential featural representation of various concepts. Intersection:Union ratio shown along line connecting concepts. A: Intact concepts. B: Slightly disturbed representations. Note that ranked ratios remain similar. C: Largely disrupted representations. Note that ranked ratios are dramatically different than in A.

Board sorting tasks require an even greater amount of semantic knowledge than naming tasks to perform at a level similar to normal controls. These tasks require participants to represent the detailed relations between a number of concepts in a spatial manner. Because this task requires that participants not only identify the concept but also compare its features against those in similar concepts, partial semantic knowledge can be better distinguished from complete semantic knowledge than in tasks such as picture pointing and picture naming. It is also worth noting that, unlike serial similarity tasks, board sorting presents patients with many items from a category of concepts all at once. Activating related concepts thus strengthens both the representation of the individual concept and the underlying structure of the category as a whole. Additionally, board sorting tasks, like serial similarity tasks, can give a different type of information, particularly that of relational knowledge (i.e., is concept A more related to concept B or concept C?). The previous studies that have used this task have taken advantage of this aspect of the task by using analyses such as the Pathfinder algorithm, which investigate relational knowledge (Bonilla & Johnson, 1995; Ober & Shenaut, 1999).

Along with the different degrees of knowledge needed to perform the various tasks, picture naming and picture pointing tasks result in qualitatively different data than board sorting tasks. The reason for this is that the picture naming and picture pointing tasks, when measuring accuracy rather than reaction time, are largely all-or-none tasks in that, if the partial semantic knowledge happens to contain the features essential for naming that picture, then naming is performed perfectly and no deficit is apparent (at least for that particular item). However, if the same amount of partial semantic knowledge consists largely of non-distinguishing features, then naming fails and a percent correct score reveals no semantic knowledge for that item. Simply averaging across participants does not eliminate the problem of naming errors not being linearly related to the amount of semantic knowledge because the nature of the deficit is largely idiosyncratic. In the board sorting task, all of the semantic knowledge available can potentially have a direct effect on the closeness of two specific chips on the board.

4.3. Similarity structure and richness of semantic information

The findings of the experiments we presented are consistent with a notion of concepts as patterns of activation across sets of distributed features, with connections among these features randomly damaged in patients with AD, resulting in semantic deficits. This suggests two things: first, these patients have deficits in featural knowledge; second, although semantic deficits may exist, this does not mean that the entire network underlying a word's meaning has been damaged.

Results from similarity tasks (especially the simultaneous similarity tasks) demonstrate a different aspect of semantic knowledge. Although there is no inherent reason why serial and simultaneous similarity judgment tasks cannot be analyzed in the same way, in the above-cited studies the simultaneous similarity studies either compared the responses from the AD group with those from the NC group in terms of chance or in terms of general dimensions used, but the serial similarity judgment studies generally examined the data at a more fine-grained level. This difference in analysis crucially explains why different results have been obtained in the serial and simultaneous tasks.

The results of the similarity judgment studies point to a number of interesting and important characteristics of the semantic deficits of these patients. First, as noted by the serial similarity judgment tasks, under close analysis, with the AD group there is less consistency in terms of which aspects of meaning the judgments are based on. This supports the notion that these patients are randomly losing featural knowledge; thus some may be able to use a certain set of features to judge the similarity of two objects, whereas for others these features may be unavailable and other features must be used to make the similarity judgments.

The results from the board sorting tasks look at a different aspect of semantic knowledge and are also consistent with an approach whereby features are randomly damaged. As Fig. 6 illustrates, as features are randomly damaged, the relative relations between the various concepts can be maintained. The loss of featural knowledge can, however, distort the similarity structure of the category, as seen with the analysis for the serial similarity judgment tasks, but this distortion is minimal early on, especially for sets of items that are not highly similar. Since the analyses used previously for the board sorting task merely examined the relations between the various concepts, it is expected that the patients appear relatively normal since the general relations may not be destroyed at low levels of damage. The finding that the patients' representations look somewhat less normal as the disease progresses is also consistent with this approach, since larger amounts of damage increase the distortion of the relations. Additionally, because the categories used in these tasks often involved few highly related items (e.g., all words did not belong to a small, tight class such as domestic animals, but rather to a bigger superset, allowing for a greater degree of difference between the concepts), one would expect the effects of the distortions that exist in the patients' semantic space to be minimized.

In light of the research cited above, an explanation emerges for the results of the current study. Although the results from the various analyses suggest both that the patients share similar semantic representations with the YN/ON group and that the patients have impaired representations, both of these results are consistent with a unitary explanation whereby damage is slowly affecting the semantic representations of the patients, making various features unavailable. This accounts for why, across all tasks and analyses, the AD group shows some deficit compared to the YN/ON group. However, since these patients are in mild to moderate rather than severe stages of AD, it is likely that many features are still available to them. This explains why these patients appear comparable but not identical to the YN/ON group when using tasks and analyses that investigate the relative similarity structure. This suggests that the result of AD is to distort the semantic space, which eventually will lead to more comprehensive disintegration of semantic knowledge, but in earlier stages results in partially intact information; the patients will have some naming problems when critical features are unavailable, but in general will know that a *dog* is more similar to a *cat* than it is to a *camel*.

As mentioned in Section 1, in addition to studies using the board sorting task, some studies using other similarity-based tasks have found that AD patients appear to show normal performance. In one such study by Rich et al. (2002), participants sorted nine animal concepts into three categories. The AD patients' grouping was statistically identical to that of normal controls. To understand how these results are consistent with the theory put forth here requires an examination of the stimuli and the task. The stimuli Rich et al. chose came from three distinct subcategories of the category animals: wild animals, pets, and zoo animals. By using items from distinct subcategories, they improved the chances that, even with damage across a number of features, those items within a subcategory would share even more features than those items across subcategories. The finding that, when allowed to use as many groups as they wanted, the AD patients created more groups than the normal controls (although the contents of the piles were semantically motivated) is also consistent with the notion of damage to features in a distributed semantic system. As features become randomly damaged, the larger structure of the category may no longer be apparent and thus less "meaningful" characteristics are used to group concepts. This shift away from reliance on features that normal controls rely on for similarity tasks has been noted by Chan and colleagues (Chan et al., 1993): "If as suggested by the present results, AD patients focus on concrete attributes like size rather than more abstract features like domesticity, AD patients' choice of intermediate and nonessential attributes on an associative-ranking task is understandable" (p. 417).

Although it may appear that our data is in conflict with that of Chan et al. (2001), who found undisrupted representations for *tools* (but not for *animals*) on a triadic comparison similarity task, their results are in fact predicted by our account. The critical difference between the category structures of *tools*, for which the representations were normal, and *animals*, for which they were not, is the density of the semantic space for each. Because natural kinds categories occupy a dense semantic space (with items sharing a large number of features), slight distortions of the semantic system will tend to have a more disruptive effect on the organization of those categories than for artifact categories. Damage to artifact categories, where items are more spread out in semantic space, will tend to have less of an impact on category organization because considerable damage must occur before one concept enters into the semantic space of another. Thus, our theory predicts that natural kinds would be more subject to damage than artifacts when relational knowledge is tested. We have independently confirmed this prediction elsewhere (Aronoff, Gonnerman, Almor, Kempler, & Andersen, 2004).

It should be noted that our results also suggest that the board sorting task can reveal important information that is not accessible through tasks such as picture naming. This task generates two types of data. The first type, analyzed using correlations and the Pathfinder algorithm, contains information about the general similarity structure of a semantic category. The second type, analyzed by the spread measure, contains information about the quantity of information contained within the concepts in a given category. Together, these two measures can provide important insight into the semantic representations of patients with AD, as well as other special populations.

Acknowledgement

This research was supported by NIA grant R01 AG-11774-04. Additionally, the first author was supported by NIH training grant 5T32MH20003-05. UnJa Hayes, Houri Hintiryan, Alison Flipse, and Mariko Taki provided invaluable help in stimuli preparation, testing, and data coding. We thank the participants and their families for their generous contribution of time and effort. We would also like to thank the two anonymous reviewers for their helpful and insightful comments. Portions of this work were presented at the annual meetings of the Academy of Aphasia (Vienna, Austria, October 2003) and the Cognitive Science Society (Boston, August 2003).

Appendix A

Stimulus list

Domestic animals (8/8/5[a]):	Duck, sheep, goose, dog, cow, horse, donkey, cat, goat, pig, chicken, camel
Zoo animals (8/8/3):	Zebra, rhinoceros, walrus, elephant, giraffe, gorilla, kangaroo, lion, monkey, penguin, leopard, tiger
Local wildlife (8/8/0):	Moose, fox, bear, mouse, rabbit, raccoon, skunk, squirrel, frog, deer, beaver, bird
Insects (8/8/9):	Mosquito, scorpion, worm, ant, bee, beetle, butterfly, caterpillar, fly, grasshopper, moth, spider
Vegetable (8/8/4):	Asparagus, cabbage, garlic, carrot, lettuce, celery, corn, potato, onion, pepper, pea, pumpkin
Fruit (8/8/8):	Pineapple, watermelon, plum, apple, banana, cherry, grapes, lemon, orange, pear, peach, strawberry
Carpenter's tools (8/8/4):	Screwdriver, ladder, drill, axe, chisel, nail, hammer, pliers, saw, screw, vise, wrench
Kitchen items (9/8/4):	Spatula, stove, mixer, pitcher, cup, broom, frying pan, kettle, pot, rolling pin, bowl, toaster
Clothing (8/8/5):	Coat, suit, apron, vest, boot, cap, necklace, glove, hat, ring, jacket, mitten
Musical instruments (8/8/8):	Xylophone, harpsichord, clarinet, accordion, drum, flute, French horn, guitar, harp, piano, trumpet, violin
Vehicles (10/8/4):	Scooter, tractor, sled, airplane, bicycle, roller-skate, helicopter, motorcycle, sailboat, canoe, ship, truck
Toys (8/8/6):	Rattle, dice, tricycle, dart, balloon, wagon, kite, swing, top, football, blocks, baseball bat

[a] The number of YN/ON/AD participants that completed this board, respectively.

References

Aronoff, J. M., Gonnerman, L. M., Almor, A., Kempler, D., & Andersen, E. S. (2004). The role of similarity structure in category specific deficits in Alzheimer's disease. *Brain and Language*, *91*(1), 154–155.

Aronoff, J. M., Gonnerman, L. M., Andersen, E. S., Kempler, D., & Almor, A. (2003). Implications of distributed representations for semantic processing: Evidence from Alzheimer's disease. *Proceedings of the Cognitive Science Society*, *25*, 97–102.

Astell, A. J., & Harley, T. A. (1998). Naming problems in dementia: Semantic or lexical? *Aphasiology*, *12*(4/5), 357–374.

Astell, A. J., & Harley, T. A. (2002). Accessing semantic knowledge in dementia: Evidence from a word definition task. *Brain and Language*, *82*(3), 312–326.

Barbarotto, R., Capitani, E., Jori, T., Laiacona, M., & Molinari, S. (1998). Picture naming and progression of Alzheimer's disease: An analysis of error types. *Neuropsychologia*, *36*(5), 397–405.

Barbarotto, R., Capitani, E., & Laiacona, M. (2001). Living musical instruments and inanimate body parts? *Neuropsychologia*, *39*, 406–414.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, *80*(3), 1–46.

Bayles, K. A., Tomoeda, C. K., & Cruz, R. F. (1999). Performance of Alzheimer's disease patients in judging word relatedness. *Journal of the International Neuropsychological Society*, *5*, 668–675.

Bonilla, J. L., & Johnson, M. K. (1995). Semantic space in Alzheimer's disease patients. *Neuropsychology*, *9*(3), 345–353.

Chan, A. S., Butters, N., & Salmon, D. P. (1997). The deterioration of semantic networks in patients with Alzheimer's disease: A cross-sectional study. *Neuropsychologia*, *35*(3), 241–248.

Chan, A. S., Butters, N., Salmon, D. P., & McGuire, K. A. (1993). Dimensionality and clustering in the semantic network of patients with Alzheimer's disease. *Psychology and Aging*, *8*(3), 411–419.

Chan, A. S., Salmon, D. P., & De La Pena, J. (2001). Abnormal semantic network for "animals" but not "tools" in patients with Alzheimer's disease. *Cortex*, *37*, 197–217.

Chertkow, H., & Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type: What do various measures measure? *Brain*, *113*, 397–417.

Davidoff, J., & Roberson, D. (2004). Preserved thematic and impaired taxonomic categorization: A case study. *Language and Cognitive Processes*, *19*(1), 137–174.

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience*, *10*(1), 77–94.

Duncan, G. T., & Layard, M. W. (1973). A Monte-Carlo study of asymptotically robust tests for correlation. *Biometrika*, *60*, 551–558.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton-Mifflin.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

Giustolisi, L., Bartolomeo, P., Daniele, A., Marra, C., & Gainotti, G. (1993). Category-specific semantic impairment for living things in the early stages of Alzheimer's disease: Further evidence from a study on single cases. *Journal of Clinical and Experimental Neuropsychology*, *15*, 403.

Gonnerman, L. M., Andersen, E. S., Devlin, L. T., Kempler, D., & Seidenberg, M. S. (1997). Double dissociation of semantic categories in Alzheimer's disease. *Brain and Language*, *57*, 254–279.

Gonnerman, L. M., Aronoff, J. M., Almor, A., Kempler, D., & Andersen, E. S. (2004). From Beetle to bug: Progression of error types in naming in Alzheimer's disease. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (p. 1563).

Hayes, S. C., & Bissett, R. T. (1998). Derived stimulus relations produced mediated and episodic priming. *Psychological Record*, *48*(4), 617–630.

Hodges, J. R., Salmon, D. P., & Butters, N. (1991). The nature of the naming deficit in alzheimers and huntingtons disease. *Brain*, *114*, 1547–1558.

Huff, F. J., Corkin, S., & Growden, J. H. (1986). Semantic impairment and anomia in Alzheimer's disease. *Brain and Language*, *28*, 235–249.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin and Review*, *10*(4), 785–813.

Kirshner, H. S., Webb, W. G., & Kelly, M. P. (1984). The naming disorder of dementia. *Neuropsychologia*, *22*, 23–30.

Laiacona, M., Barbarotto, R., & Capitani, E. (1998). Semantic category dissociations in naming: Is there a gender effect in Alzheimer's disease? *Neuropsychologia*, *36*(5), 407–419.

Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin and Review*, *7*, 618–630.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.

Nakamura, H., Nakanishi, M., Hamanaka, T., Nakaaki, S., & Yoshida, S. (2000). Semantic priming in patients with Alzheimer and semantic dementia. *Cortex*, *36*, 151–162.

Nebes, R. D. (1989). Semantic memory in Alzheimer's disease. *Psychological Bulletin*, *106*(3), 377–394.

Ober, B. A., & Shenaut, G. K. (1999). Well-organized conceptual domains in Alzheimer's disease. *Journal of the International Neuropsychological Society*, *5*(7), 676–684.

Paganelli, F., Vigliocco, G., Vinson, D., Siri, S., & Cappa, S. (2003). An investigation of semantic errors in unimpaired and Alzheimer's speakers of Italian. *Cortex*, *39*, 419–439.

Rich, J. B., Park, N. W., Dopkins, S., & Brandt, J. (2002). What do Alzheimer's disease patients know about animals? It depends on task structure and presentation format. *Journal of the International Neuropsychological Society*, *8*, 83–94.

Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorization: Neuropsychological evidence for a dissociation in explicit categorization tasks. *Cognition*, *71*, 1–42.

Rom, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika*, *77*, 663–666.

Sacchett, C., & Humphreys, G. W. (1992). Calling a squirrel a squirrel but a canoe a wigwam: A category-specific deficit for artifactual objects and body parts. *Cognitive Neuropsychology*, *9*(1), 73–86.

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization*. Norwood, NJ: Ablex.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, England: Cambridge University Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

Snow, C. (1990). The development of definitional skill. *Journal of Child Language*, *17*, 697–710.

Somners, L. M., & Pierce, R. S. (1990). Naming and semantic judgements in dementia of the Alzheimer's type. *Aphasiology*, *4*, 573–586.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Vinson, D. P., Vigliocco, G., Cappa, S., & Siri, S. (2003). The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language*, *86*, 347–365.

Warrington, E. K., & Shallice, T. (1979). Semantic access dyslexia. *Brain*, *102*, 43–63.

Wilcox, R. R. (2003). *Applying Contemporary Statistical Techniques*. San Diego, California: Academic Press.

Wilcox, R. R., & Muska, J. (2002). Comparing correlation coefficients. *Communications in statistics: Simulation and Computation*, *31*(1), 49–59.

Williams, J. N. (1996). Is automatic priming semantic? *European Journal of Cognitive Psychology*, *22*, 139–151.

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behavioral Research Methods. *Instruments and Computers*, *20*, 6–11.

Yu, M. C., & Dunn, O. J. (1982). Robust tests for the equality of two correlations: A Monte Carlo study. *Educational and Psychological Measurement*, *42*, 987–1004.