Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 99 (2006) 289–303
Brain and Language
www.elsevier.com/locate/b&l

Word association in early Alzheimer's disease ☆

Tamar H. Gollan [a,*], David P. Salmon [b], Jessica L. Paxton [c]

[a] *Department of Psychiatry, University of California, San Diego, USA*
[b] *Department of Neurosciences, University of California, San Diego, USA*
[c] *Department of Psychology, Washington University, St. Louis, MO, USA*

Accepted 6 July 2005
Available online 24 August 2005

Abstract

The hypothesis that Alzheimer's disease (AD) degrades semantic representations predicts that AD qualitatively alters spontaneous thoughts. In two experiments contrasting free associations to words with strong (e.g., *bride-groom*) versus weak (e.g., *body-leg*) associates participants with AD produced less common responses (e.g., *bride-pretty*) than normal controls but only for words with strong associations, and only on the first (but not on second or third) association response. Furthermore, all participants produced fewer semantically related responses to words with weak associates. Because strong associations should be retrieved more easily than weak associations these results are problematic for retrieval-based accounts of AD. Instead we propose that AD entails a semantic deficit, and that strong associations involve more semantic processing than weak associations (in all speakers).

Keywords: Word-association; Free-association; Production; Speakers; Dementia; Aging

1. Introduction

Alzheimer's is a common degenerative disease that affects the brain and from which there is no recovery. People with Alzheimer's gradually lose more and more of the cognitive abilities that normal adults take for granted ultimately including even the most basic skills such as understanding the meanings of frequent words. Although the hallmark of Alzheimer's disease (AD) is impaired explicit memory, patients with early-stage disease also display a number of symptoms consistent with the hypothesis that AD involves a gradual deterioration of semantic memory (for review, see Chertkow & Bub, 1990; Martin, 1992; Salmon, Butters, & Chan, 1999). This *degraded semantics* hypothesis is supported by data from verbal fluency tasks which showed that patients with AD were more impaired relative to healthy elderly controls on semantic fluency than on phonemic fluency tasks (e.g., patients with AD had more difficulty listing members of the category *Animals* than with listing words that begin with the letter *F*; e.g., Butters, Granholm, Salmon, Grant, & Wolfe, 1987). Semantic fluency also declined faster than phonemic fluency as the disease progressed (Salmon, Heindel, & Lange, 1999). In addition, subcategory cues (e.g., *pets, farm animals*) did not improve semantic fluency scores in patients with AD (although such cues did improve fluency scores in demented patients with Huntington's or Parkinson's disease; Randolph, Braun, Goldberg, & Chase, 1993). Impressively, patients with AD had faster mean response latencies than normal controls during semantic fluency (despite producing fewer exemplars) but not during phonemic fluency—a pattern suggesting that there is a

☆ This research was supported by a Career Development Award from NIDCD DC00191 to Tamar H. Gollan, and by NIA Grant R01 AG12963 to David P. Salmon, and NIA Grant P50 AG05131 to the University of California. We thank David Balota for valuable comments on an earlier version of this manuscript, and Cynthia Cera, Efrat Stark, and Rosa Montoya, for assistance with data coding.
* Corresponding author.
E-mail address: tgollan@ucsd.edu (T.H. Gollan).

0093-934X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2005.07.001

reduction in the number of items available for retrieval from semantic memory (Rohrer, Wixted, Salmon, & Butters, 1995). Taken together, these findings suggest that patients with AD suffer a degradation of the concepts and associations that allow cognitively intact people to efficiently produce exemplars from a semantic category, but relatively less degradation of the lexicon and associations between word forms that are important for phonemic fluency.

The degraded semantics hypothesis predicts that patients with AD should perform differently from controls on the word-association task. Word association is one of the oldest psychological tasks, dating back to Freud and Jung who suggested that word associations would reveal censored and unconscious aspects of people's thoughts. Recent interpretations of the word association task also stipulate that meaning drives responses (de Groot, 1989) since the majority of responses generated in word-association are related to the stimulus word in meaning. For example, given *cat* people will be much more likely to say *dog* than *can* or *bat* (Postman & Keppel, 1970). A compelling reason for postulating that meaning (rather than associations between word-forms) drives responses in word association comes from research on bilingual speakers. When asked to free associate to a stimulus word either within or across languages, bilinguals were almost as likely to produce the same associated concept across languages as they were within the same language. That is, given the English word *cat* a Spanish–English bilingual would be almost as likely to produce *perro*, the Spanish word for *dog*, during a between-language testing condition as she was to produce *dog* during a within-language condition (van Hell & de Groot, 1998). Word-forms in different languages are extremely unlikely to be related associatively (because they rarely occur together) therefore this finding strongly suggests that word associations primarily reflect the characteristics of stored meanings (which would be shared between languages).

In accord with the degraded semantics hypothesis, patients with AD had particular difficulty with the word association task; in many cases, patients were able to provide meaningful stimulus definitions but could not provide a single associate (Abeysinghe, Bayles, & Trosset, 1990). Instead participants with AD produced multi-word responses, repeated the stimulus, or provided unrelated responses. In addition, they were significantly impaired in discriminating highly related from unrelated associates, and were less likely to provide an associate in the same grammatical class relative to controls (for similar results in dementia in general see Gewirth, Shindler, & Hier, 1984; Santo Pietro & Goldfarb, 1985). These results clearly demonstrated that patients with AD perform differently than cognitively intact controls on the word association task, but did not clarify precisely how AD altered the types of associations produced.

To understand more about the nature of cognitive impairment in AD, it may be informative to examine the types of semantic associations that patients with AD produce (when they do produce them). Two studies utilized this approach. In one study, dementia severity correlated with response commonality for both English and French speakers with AD; as cognitive impairment increased, patients were less likely to give common responses (Eustache, Cox, Brandt, & Lechevalier, 1990). In a second study, Randolph (1991) demonstrated that patients with AD produced less typical responses on the word-association task than did age-matched controls, and this difference appeared to be numerically larger when stimulus words had strong associates than when they had weak associates. Unfortunately, Randolph was unable to explore this apparent relationship because the interaction between participant and stimulus type was not significant. The failure to observe a significant interaction may have been due to insufficient power given the small sample size or to a relatively small difference in associative strength between so-called strong and weak associates in the Randolph study. In any event, both of these studies suggest that strong associates are more vulnerable than weak associates to the deleterious effects of AD, but given the studies' limitations, further analysis of the role of associative strength on the word association responses of patients with AD is warranted.

The current study was designed to better characterize the nature of word association responses in AD to constrain accounts of the cognitive impairments associated with the disease, and accounts of the word association task. The goal of the first experiment was to determine if stimuli with strong associations (i.e., *strong stimuli*) in fact produce greater differences between patients with AD and age-matched controls relative to stimuli with weak associations (i.e., *weak stimuli*). An example of a stimulus with a very strong association is *dill*, to which 87% of cognitively intact participants produced *pickle* (Nelson, McEvoy, & Schreiber, 1998). In contrast, a stimulus with weak associates is *chicken*, to which only 9% produced its strongest associate *soup*. By comparing word-association responses to strong and weak stimuli in patients with AD, the present study evaluated whether or not associative strength is an important factor in producing the difference between patients with AD and controls on the word-association task.

The second experiment was designed to provide evidence against the possibility that a confounding factor produced the pattern of results obtained in Experiment 1. Strong and weak stimuli vary on many factors. For instance, relative to weak stimuli, strong stimuli are more likely to be concrete, and are less likely to be associated with many different words. To illustrate, after *pickle* only two additional associates to *dill* are listed in the Nelson et al. (1998) norms. These are *weed* and *spice* and both were very weakly associated to *dill* (only 1% of the participants in the normative study produced them). In Experiment 2 we tested if *dills* could be turned into *chickens* by using a continuous association task. Participants produced three associations (instead of just one) to the same stimulus. If associative strength (not concreteness or some other confounding factor) is the critical factor in producing the difference between patients with AD and controls, then even strong stimuli should produce no differences between groups after the first trial because strong associates are no longer available.

2. Experiment 1—Strong versus weak stimuli

In a previous study that examined the impact of associative strength on word association performance in AD, Randolph (1991) used only 10 strong and 10 weak items, and tested only 10 participants with AD. In addition, he used "strong" stimuli with associates that (on average) only 50% of a normative sample produced and "weak" stimuli with associates that (on average) as many as 30% of the normative sample produced. As such, his failure to observe an interaction between strength and group (AD versus control) may have reflected a Type II error. To test whether AD affects strong but not weak associations, Experiment 1 incorporated a stronger manipulation of associative strength.

2.1. Method

2.1.1. Subjects

Eighteen patients with probable Alzheimer's disease (AD) and 18 elderly normal controls (NC) participated in the study. All were participants in the University of California, San Diego (UCSD) Alzheimer's Disease Research Center (ADRC) through which they received annual medical, neurological, and neuropsychological evaluations. Based on these evaluations and a number of laboratory tests (to rule out other causes of dementia) the diagnosis of probable AD was made by two senior staff neurologists at the ADRC according to criteria developed by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) (McKhann et al., 1984). Diagnosing neurologists were provided with a general statement concerning results of the neuropsychological evaluation (e.g., a deficit in two or more areas of cognition) but were not aware of any specific test scores. To reduce the inclusion of patients with multi-infarct dementia, patients with scores of five or greater on the Rosen-modified Hachinski ischemia scale (Hachinski et al., 1975; Rosen, Terry, Fuld, Katzman, & Peck, 1980) were excluded from the AD group. The NC participants were either spouses of the patients with AD or were recruited through newspaper advertisements. Volunteers with a history of severe head injury, learning disability, alcoholism, drug abuse, or serious psychiatric disturbance were excluded. All participants or (when appropriate) caregivers provided informed consent after the study procedure had been fully explained. On average patients with AD ($M = 4.17$, $SD = 2.92$) and normal controls ($M = 4.44$, $SD = 3.68$) participated in the experiment approximately 4 months after their annual evaluation at the ADRC and the delay between the two test dates was never longer than 11 months. Table 1 shows the mean age, years of education, mean Mini-Mental State Examination scores (MMSE; Folstein, M.F., Folstein, S.E., & McHugh, 1975) and Dementia Rating Scale scores (DRS; Mattis, 1988) for each group. The AD and NC groups did not differ significantly in mean age or years of education (both ts < 1).

2.1.2. Materials

The stimulus materials for the current investigation, their division into strong versus weak categories, as well as coding of association strength of participant's responses were based on the normative database published by Nelson et al. (1998). The Nelson et al., database lists 5019 stimulus words and the responses provided by 6000 participants in a free association task in which subjects were told to "… write the first word that comes to mind that is meaningfully related or strongly associated to the word." Response strength in the current study was rated using the proportion of subjects who produced that same response in the Nelson et al., norms. Further details regarding the Nelson et al., database are available on the web at: http://cyber.acomp.usf.edu/FreeAssociation/Intro.html.

Stimuli with very strong ($n = 26$) and very weak associates ($n = 26$) were selected from the Nelson et al. (1998) norms to differ as much as possible in association

Table 1
The mean age, years of education, mini-mental state examination scores (MMSE), and dementia rating scale (DRS) scores of the normal control (NC) subjects and patients with probable Alzheimer's disease (AD) in Experiments 1 and 2

	Experiment 1		Experiment 2	
	Normal controls	Alzheimer's patients	Normal controls	Alzheimer's patients
Age	76.67 (7.79)	76.33 (6.44)	77.54 (7.33)	74.00 (13.23)
Education	16.00 (2.74)	16.22 (2.90)	15.62 (2.84)	15.08 (2.50)
MMSE	29.56 (0.86)	24.39 (2.62)	28.85 (1.28)	22.83 (3.83)
DRS	140.61 (2.20)	120.00 (9.49)	139.00 (3.44)	117.92 (8.22)

Standard deviations are shown in parentheses.

strength (called FSG in the association norms) with an attempt to avoid extremely rare words, ambiguous words (e.g., *bark*), words with an obvious antonym (e.g., *black*) or synonym (e.g., *couch*), and words that were strongly associated with a common category label (e.g., *kiwi-fruit*),[1] or were are part of compound words or phrases (e.g., *racket ball*). On average strong stimuli had an associate that was produced by 69% ($SD = 12\%$) of the normative sample, whereas on average weak stimuli had associates produced by only 9% ($SD = 3\%$) of the normative sample. Strong and weak stimuli also varied on a number of other dimensions. Specifically, strong stimuli (6.0 $SD = 1.6$) tended on average to have fewer words associated with them than weak stimuli (26.9 $SD = 5.3$; this variable was called QSS in the association norms), strong stimuli tended to be lower in word frequency (20.2 $SD = 36.5$) than weak stimuli (51.6 $SD = 78.6$; according to the CELEX database Baayen, Piepenbrock, & Gulikers, 1995), strong stimuli tended to be shorter in number of syllables (1.7 $SD = 0.8$) than weak stimuli (2.1 $SD = 0.7$), and more strong stimuli ($n = 18$ or 69%) than weak stimuli ($n = 6$ or 23%) were concrete. We defer discussion of these largely unavoidable confounds until Experiment 2 (which was designed to address this problem). A list of the materials we used in Experiments 1 and 2 is shown in the Appendix.

2.1.3. Procedure

The participants were tested individually in a quiet room. During the testing session, an experimenter verbally presented each stimulus word to each participant. Before beginning the task, participants were given the following instructions: "In this task I will say a word and then I want you to tell me the first word that comes to mind. Just say whatever word you think of after I say my word. For example, if I say *television* you might say *radio*. Or if I say *fortune* you might say *teller*. Do you have any questions?" If the participant inquired as to whether the amount of time he or she took to respond was measured, the experimenter clarified that time was not important. If the participant did not respond when given a stimulus word, the experimenter prompted the participant by saying, "What does *television* make you think of?" If the participant still didn't respond, the experimenter said, "Just say whatever comes to mind—any word is fine as long as you haven't said it already." If the participant still did not respond the experimenter said: "Let me give you another example. If I say *lobster*—that might make you think of *shrimp*." Also, if the participant responded by saying that they were unable to think of a word, the experimenter encouraged the participant once by saying "take your time and try to think of any word that comes to mind." When the participant responded with a lengthy tangential phrase or a definition of the stimulus word, the experimenter said: "try to respond with just one word that comes to mind after I say my word." If the participant produced the same association on successive trials the experimenter said "think of a different word, you have to produce a different word each time." The experimenter wrote down each response on a test form, and responses were also audio taped for subsequent verification of manual coding.

2.2. Results and discussion

We determined the associative strength of each item produced on the word association task by each participant by using the Nelson et al. (1998) norms (using values in the variable called "FSG"). When participants produced more than one word (a multi-word response; see Table 3), the most strongly associated word produced was scored. Similarly, when participants produced a word that was morphologically related to the most strongly associated word, they were credited for producing the strongest associate. For example, a participant said "cutting paper" in response to the stimulus *scissors*. This response was scored 0.87 because 87% of the normative sample produced *cut* in response to *scissors*. Responses that did not appear in the Nelson et al. (1998) norms were coded as zero. We calculated the means and standard deviations of association response strength for the two groups in each condition and these are shown in Table 2. We calculated separate ANOVAs for subject means ($F1$) and stimulus item means ($F2$). The analyses included two factors with two levels: Group Type (AD and NC) and Stimulus Strength (strong and weak). Group type was always treated as a between subjects variable, and Stimulus Strength was a repeated factor in the subjects analysis and a between items factor in the items analysis.

The results of the subjects analysis are shown in Table 2. There was a robust main effect of stimulus type such that participants produced more common responses when they were given a strong stimulus $F1(1,34) = 537.82$,

Table 2
Participants' mean response strength (from Nelson et al., 1998) in Experiment 1

Stimulus type	Group type	
	AD	NC
Strong	0.33 (0.16)	0.41 (0.18)
Weak	0.02 (0.01)	0.02 (0.01)

Standard deviations are shown in parentheses.

[1] Although we presented it first, Experiment 1 was actually run after Experiment 2. The category restriction was not employed in Experiment 2. For example, one of the stimuli in Experiment 2 was the stimulus *cat* that is a member of the frequently encountered semantic category *animal*. Because different mechanisms may be involved in generating category labels instead of other types of associations we decided to avoid such stimuli in Experiment 1.

$MSE = 0.004$, $p < .01$; $F2(1,100) = 230.22$, $MSE = 0.014$, $p < .01$. There was also a main effect of group type that was significant in the subjects analysis $F1(1,34) = 6.46$, $MSE = 0.004$, $p < .05$ and marginally significant in the items analysis $F2(1,100) = 3.18$, $MSE = 0.014$, $p = .078$. In addition, there was an interaction between group and stimulus type such that, relative to weak stimuli, strong stimuli produced a greater difference between patients with AD and the NC group. The interaction was significant in the subjects analysis $F1(1,34) = 6.97$ $MSE = 0.004$, $p = .01$ and there was a trend in the same direction in the items analysis $F2(1,100) = 2.98$, $MSE = 0.014$, $p = .087$.

The main effect of group type supports the degraded semantics hypothesis that predicted that patients with AD should produce different types of semantic associations compared to NC participants. Two aspects of the results require further explanation. First, patients with AD produced less common responses than NC participants, and second, AD had a greater effect on strong associations than on weak associations. In fact, planned comparisons indicated that AD had no effect on responses to weak stimuli ($t < 1$). Before considering the implications of these results we conducted some additional analyses to provide further evidence that semantic factors produced the observed findings.

2.3. Subjective analysis

To further assess the nature of the differences between groups a research assistant (who was blind to the hypothesis being tested) coded each response (to all the stimuli in both strong and weak conditions) into one of the following categories: semantic (e.g., *scissors*-*paper*), form (e.g., *scissors*-*sauces*), both (e.g., *scissors*-*sew*), multi-word (e.g., *scissors*-*cut paper*), unrelated (e.g., *scissors*-*quiet*), nonword (e.g., *lizard*-*skirk*; the stimulus *lizard* was one of few that led to a nonword response; no nonwords were produced in response to the stimulus *scissors*). The results of this analysis are shown in Table 3. The majority of responses were semantically related to the stimulus. This is consistent with previous suggestions (see Section 1) that the word association task taps semantic processing.

Unrelated responses were the next most common response type (after semantically related responses) in the NC group, and the third most common response type in the AD group. It is possible, however, that these responses reflected idiosyncratic semantic relationships that were not obvious without explanation. For example, in response to the stimulus *scissors* one participant said *never*; this response seems unrelated to *scissors*, however, it is possible that this person was thinking "I never use scissors." In the AD group multi-word responses were second most common (after semantically related responses). In fact, participants in the AD group generated significantly more multi-word responses than did NC participants. The same result was reported previously (Abeysinghe et al., 1990) and may reflect the AD group's greater difficulty with following the task instructions relative to the NC group. A number of other marginally significant differences emerged between participant groups (see Table 3); we defer discussing these until after presenting Experiment 2.

2.3.1. Semantically related versus not semantically related responses

One reason why participants in the AD group may have produced less common responses on average would be if they produced more odd responses because they temporarily lost track of the instructions or forgot the stimulus. Odd responses would be scored a zero and this would reduce the semantic association scores of AD participants on average without there having been any actual qualitative change in the type of semantic responses that the AD group produced. If so, the reduction in response commonality could be explained without relying on the degraded semantics hypothesis.

To address this possibility we compared the average association response strength (again using the variable FSG in the Nelson et al., 1998, norms) in the two

Table 3
The mean percentage of responses classified into different types in Experiments 1 and 2 collapsed across first, second, and third association trials

Experiment number	Group type or statistic	Response type						
		Semantic	Form	Both[a]	Multi-word	Unrelated	Nonword	No response
Experiment 1	NC	80.1 (12.8)	0.4 (2.4)	5.1 (2.1)	6.2 (7.4)	7.8 (5.5)	0.2 (0.7)	0.1 (1.5)
Experiment 1	AD	74.8 (20.9)	1.3 (2.4)	3.6 (2.1)	11.5 (7.4)	7.6 (5.5)	0.3 (0.7)	0.9 (1.5)
	Two-tailed[b] *t* test	1.66	1.44	1.49	2.54	<1	<1	2.02
	p value	0.11	0.16	0.15	0.02	0.90	0.64	0.05
Experiment 2	NC	80.8 (7.7)	1.3 (1.9)	3.1 (1.6)	4.4 (4.9)	10.0 (4.8)	0.4 (0.8)	0.0 (0.0)
Experiment 2	AD	73.9 (20.9)	1.1 (2.7)	0.9 (1.8)	15.0 (14.4)	7.3 (18.1)	0.4 (0.0)	1.9 (4.4)
	Two-tailed[b] *t* test	1.47	<1	3.21	2.46	1.39	1.82	1.45
	p value	0.16	0.81	<0.01	0.02	0.18	0.08	0.16

Standard deviations are shown in parentheses.
[a] Responses that were both semantically and formally related to the stimulus word.
[b] Degrees of freedom = 34 in Experiment 1 and = 23 in Experiment 2.

participant groups including only responses that were judged as semantically related to the stimulus in the strong-stimuli condition (the condition that demonstrated a significant difference between participant groups). The degraded semantics hypothesis predicts that patients with AD should produce less common responses even when limiting the comparison exclusively to responses that were semantically related (and therefore were not "odd"). The results of this analysis confirmed this prediction.

To the extent that AD participants did produce semantically related responses, these were still less common responses on average 0.38 ($SD = 0.11$) than those that participants in the NC group 0.46 ($SD = 0.09$) produced. This difference was significant $t(34) = 2.24$, $p = .032$. This finding is of interest because it suggests that the reduction in response commonality in the AD group (for strong stimuli; see Table 2) was not simply an artifact of between group differences in their approach to, or their ability to carry out, the word-association task. Even when participants with AD produced words that were semantically related to the stimulus words the responses tended to be less common than those that the NC group produced.

3. Experiment 2—Continuous association

The results of Experiment 1 provided evidence that strong associations are more vulnerable than weak associations to the cognitive impairments associated with AD. The materials in Experiment 1 entailed the strongest possible manipulation of association strength. As noted (see Section 2.1.2) this resulted in the inadvertent manipulation of a number of other variables as well (e.g., concreteness, number of associates). To illustrate, a typical strong stimulus is the word *shingle*. It is concrete, and was strongly associated with the word *roof* (61% of responses in the Nelson et al., 1998, norms) and weakly associated with only eight other words in the norms (10% with *house*, and 1% with *crystal, disease, money, tile, wall*, and *window*). In contrast, a typical weak stimulus is the word *crisis*. It is abstract, it's strongest associate is *problem* which was produced by only 10% of the normative sample, and it was also associated with 29 other words in the norms (7% with *emergency*; 5% with *trouble*; 3% with *tragedy, disaster, dilemma, danger, stress, help*; 2% with *serious, sad, panic, bad*; and 1% with *worry, trauma, situation, school, oil, missile, mid-life, hurt, handle, emotion, death, cry, crash, catastrophe, attack*, and *anxiety*).

When considering association strength, after the first and strongest associate is exhausted, strong stimuli essentially become weak stimuli. For example, the second strongest associate to strong stimulus *shingle* is the word *house* that was produced by 10% of the normative sample much like the first associate to the weak stimulus *crisis* (i.e., 10% *problem*). If association strength is the variable responsible for the difference between groups, rather than one of the other variables confounded with stimulus strength, then the difference should go away after participants generated the first association. To test this prediction in Experiment 2 we asked participants to generate three successive word associations to each stimulus, on average the strongest possible association to the stimuli in Experiment 2 was 41% ($SD = 32\%$) the second strongest was 11% ($SD = 8\%$) and the third was 9% ($SD = 6\%$). We predicted that strong stimuli would produce a difference between groups (as in Experiment 1) but only on the first response and not for subsequent responses that only had relatively weak association strength. Experiment 2 also included stimuli with a range of stimulus strengths thus addressing another possible concern regarding the results of Experiment 1 that weak stimuli did not produce a difference between participant types because of a floor effect.

3.1. Method

3.1.1. Subjects

Fourteen patients with probable Alzheimer's disease (AD) and 14 elderly normal controls (NC) participated in the study. They were from the same population as reported in Experiment 1. Two of the participants in the AD group defined the stimulus words instead of producing associations, and one of the participants in the NC group produced a majority of responses that seemed to be completely unrelated to the stimulus words. For example, in response to *gymnastics* this NC participant said *dead*, and in response to *cat* he said *afterwards* (68% of his responses were unrelated; a proportion that was 12*SD* greater than average for the NC group as shown in Table 6). These participants' data (2 from the AD group and the 1 NC participant just described) were not included in any of the analyses reported below.

Table 1 shows the mean age, years of education, mean Mini-Mental State Examination scores (MMSE; Folstein et al., 1975) and Dementia Rating Scale scores (DRS; Mattis, 1988) for each group. The AD and NC groups did not differ significantly in mean age or years of education (both ts < 1). All participants or (when appropriate) caregivers provided informed consent after the study procedures were fully explained. On average patients with AD ($M = 5.75$, $SD = 4.14$) and normal controls ($M = 3.00$, $SD = 3.58$) participated in the experiment approximately 4 months after their annual evaluation at the ADRC and the delay between the two test dates was never longer than 11 months.

3.1.2. Materials

Stimuli with very strong ($n = 6$), medium ($n = 6$), and very weak associates ($n = 6$) were selected from the Nelson et al. (1998) norms. The stimuli were selected as

Table 4
Participants' mean response strength (from Nelson et al., 1998) in Experiment 2

Stimulus type	First association		Second association		Third association	
	Group type		Group type		Group type	
	AD	NC	AD	NC	AD	NC
Strong	0.45 (0.17)	0.59 (0.11)	0.05 (0.09)	0.08 (0.09)	0.07 (0.09)	0.01 (0.04)
Medium	0.16 (.06)	0.14 (.05)	0.07 (.02)	0.08 (.03)	0.03 (.03)	0.06 (.05)
Weak	0.03 (0.02)	0.04 (0.01)	0.03 (0.02)	0.02 (0.02)	0.01 (0.01)	0.02 (0.01)

Standard deviations are shown in parentheses.

described in Experiment 1 except that an effort was made to include words with moderately strong associates in addition to the strong and weak stimuli. On average strong stimuli had an associate that was produced by 82% ($SD = 6\%$) of the normative sample, medium stimuli had an associate that was produced by 30% ($SD = 14\%$) of the normative sample, and weak stimuli had associates produced by only 12% ($SD = 2\%$) of the normative sample (see Nelson, McEvoy, & Dennis, 2000 for a suggestion that values above 4% will tend to be off the floor).

3.1.3. Procedure

The procedure was identical to that in Experiment 1 except for the instructions that were as follows: "In this task I will say a word three times. Each time I want you to try to think of a different word that comes to mind. Just say whatever word you think of after I say my word. For example, if I say *television* you might say *radio*, then I'll say *television* again and this time you might say *show*, then I'll say *television* one more time and you might say *watch*. Try to think of a different word each time. Do you have any questions?" If the participant repeated the same association then the experimenter said, "think of a different word, you have to produce a different word each time." If, after responding to the first and/or second trial, the participant stated that he or she was unable to think of any other words, the experimenter encouraged the participant once by saying "take your time and try to think of any other word that comes to mind."

3.2. Results and discussion

Means and standard deviations of association response strength were calculated and coded as in Experiment 1 for each participant and for each item in each condition. Separate ANOVAs were calculated for participant (i.e., subject) means ($F1$) and stimulus item means ($F2$). The analyses included three factors: Group Type (AD and NC), Stimulus Strength (strong, medium, and weak), and Association Trial (first, second, and third). Group Type was always treated as a between subjects variable, Stimulus Strength was a repeated factor in the subjects analysis and a between items factor in the items analysis, and Association Trial was a repeated factor in both subjects and items analyses.

The results of the subjects analysis are shown in Table 4. There were main effects of group type, stimulus strength, and association trial. Participants with AD produced less associated responses, a difference that was significant by subjects $F1\,(1,23) = 6.65$, $MSE = 0.002$, $p = .017$ but not by items $F2\,(1,30) = 1.07$, $MSE = 0.008$, $p = .309$ suggesting that only some items (in this case the strong stimuli; see below) were showing a difference between the AD and NC participants. In contrast, the main effects of stimulus strength $F1\,(1,23) = 658.09$, $MSE = 0.002$, $p < .01$; $F2\,(1,30) = 37.65$ $MSE = 0.008$, $p < .01$ and association trial $F1\,(1,23) = 239.34$, $MSE = 0.006$, $p < .01$; $F2\,(1,30) = 64.90$ $MSE = 0.005$, $p < .01$ were robust in all analyses. Both group types produced more common responses when given stimuli with stronger associates, and after the first trial their responses became less common.

There were a number of significant two-way interactions which will not be reported because they were part of a three-way Group Type × Stimulus Type × Association trial interaction that was robust by subjects $F1\,(1,23) = 8.34$, $MSE = 0.007$, $p = .008$ with a trend in the same direction in the analysis by items $F2\,(1,30) = 2.13$, $MSE = 0.005$, $p = .136$. Visual inspection of Table 4 clarifies the nature of this interaction; the between group difference was restricted almost exclusively to the first association trial in the strong condition.

These results confirm that high association strength is most sensitive to the cognitive changes associated with AD, and indicate that this finding is not attributable to a different property correlated with words that have strong associates. Words with strong associates produced a difference between group types only on the first association; the same stimuli produced no difference between groups on second and third association trials.

The results for medium strength stimuli in Experiment 2 also provided evidence against the possibility that a floor effect (in responses to weak stimuli) produced the interaction between stimulus strength and participant type (in both Experiments 1 and 2). As shown in Table 4, medium strength stimuli led to responses that were more typical than weak stimuli (i.e., responses that were off the floor), however, there were still no significant differences between groups ($t < 1$; and in fact, the difference was in the wrong direction with AD participants producing more typical responses than NC participants). We provide additional evidence against floor effects in Section 4.

A possible criticism of the method in Experiment 2 is that after producing the first response participants could have produced associations to their own first responses instead of to the stimulus (i.e., response chaining). If so, then Experiment 2 would not be convincing as evidence against the possibility that some factor correlated with stimulus strength, rather than strength itself, produced the observed findings. To assess the response chaining possibility we randomly selected 20 data points from the NC group and 20 from the AD group. If the participants were chaining their responses then their second responses should be associatively related to their own first responses. It may also seem that the second responses should be more related to the stimulus than to the first responses. However, responses cannot be more strongly related than the stimuli allow and only weaker associations remain after the strongest associations are produced. For example, given *weapon* 59% of the participants in the normative sample produced *gun*, but only 8% of the normative sample produced the second strongest association *knife*. In addition, first responses varied also in stimulus strength, a factor that would clearly affect how strongly related second responses could be to first responses. Taking *gun* (instead of *weapon*) as a stimulus, the strongest possible association was the word *shoot* produced by 24% of the participants in the sample. Thus, in comparing the second responses to the first responses we considered the baseline possible associative relatedness and expressed the relationships as a proportion.

There was no evidence to support response chaining. In the NC group, the 20 randomly chosen data points had a baseline stimulus strength (FSG) of 38% ($SD = 32\%$). The first responses to these 20 stimuli were produced by 27% ($SD = 34\%$) of the participants in the normative sample (Nelson et al., 1998) for a response typicality of about 71% (i.e., 27/38) of what could be expected relative to the younger adults in the normative sample. In contrast, treating the first responses as stimuli, these had a baseline stimulus strength of 31% ($SD = 18\%$), however the second responses were produced by only 7% ($SD = 19\%$) of the normative sample, for a response typicality of only 23% (i.e., 7/31) of what could have been expected. The random data points in the AD group provided less convincing evidence against response chaining, however, they certainly provided no positive evidence for chaining either. Response typicality was about 24% (11/41) of what could be expected on the first response and, using the first response as a stimulus, response typicality was about 20% (i.e., 7/35) of what could be expected. Positive evidence of response chaining should have shown at least a trend in the opposite direction.[2]

Table 5
The mean percentage of responses classified as semantically related to the stimulus on first, second, and third association trials in Experiment 2

	First	Second	Third
NC	91.0 (8.0)	79.5 (12.1)	71.8 (11.9)
AD	82.9 (20.2)	73.6 (21.2)	65.3 (19.3)

Standard deviations are shown in parentheses.

3.3. Subjective analysis

As in Experiment 1, subjective ratings showed that the majority of responses were semantically related to the stimuli (see Table 3). Similarly, in both experiments unrelated responses were the next most common response type (after semantically related responses) in the NC group, and the third most common response type in the AD group, and participants in the AD group generated significantly more multi-word responses relative to NC participants. A number of other significant or marginally significant between group differences emerged (see Table 3). For example, in Experiment 1 participants in the AD group were more likely to produce a "don't know" response, and in Experiment 2 NC participants were more likely to produce an associate that was related to the stimulus in both meaning and form. However, given the very low total number of responses of these types, and the lack of consistency in observed effects across experiments we do not interpret these differences any further.

The subjective analysis provided an opportunity to assess the nature of the continuous association technique, and ultimately provided a clue as to why weak stimuli were less sensitive to the effects of AD than strong stimuli (in both Experiments 1 and 2). First, as shown in Table 5, the majority of responses produced, even on the second and third association trials, were semantically related to the stimulus words. This finding is of interest because it suggests that participants approached the association task in a qualitatively similar manner on the second and third association trials, and that the reduction in differences

[2] In some cases the strongest association to the first response would have been the stimulus itself. For example, in one randomly selected data point the stimulus *winter* elicited the response *summer*. The strongest association to *summer* in the Nelson et al. (1998) norms is *winter* (produced by 40% of the sample). However, because participants were not allowed to produce the stimulus itself whenever this occurred we used the next strongest possible response (which in this case was *hot* produced by 15% of the normative sample). We made this adjustment on 3 data points in the normal control random sample, and on 1 data point in the AD sample. Without this more conservative approach the results against the response chaining hypothesis would have been even stronger because it would have increased the baseline possible similarity of the first responses as stimuli.

between groups on the second and third trials in Experiment 2 could not be attributed to a shift in how the participants approached the task on these trials.

Interestingly, although the majority of responses were semantically related, the proportion of semantically related responses decreased on average with each successive association trial (see Table 5). This finding suggested the possibility that when strong semantic associations are not available the word association task activates semantic representations to a lesser extent than when strong associations are available. Consistent with this proposal, an analysis comparing responses to weak versus strong stimuli showed fewer semantically related responses to weak stimuli. This result was robust even including only the 6 strongest and 6 weakest items in Experiment 2. On the first association trial participants produced semantically related responses 92% of the time on average ($SD = 13\%$) when given a strong stimulus, and only 72% ($SD = 21\%$) when given a weak stimulus *paired*-$t(48) = 4.22$, $p < .01$. The data from Experiment 1 demonstrated the same pattern. Responses were judged to be semantically related to the stimulus in 83% ($SD = 9\%$) of cases for strong-stimuli but only 72% ($SD = 13\%$) of cases for weak-stimuli *paired*-$t(70) = 6.35$, $p < .01$. This result suggests that weak stimuli are less likely to activate semantic representations than are strong stimuli. This proposal in turn may explain why only strong-stimuli were sensitive to AD (see below).

4. General discussion

In two experiments, patients with Alzheimer's disease (AD) and cognitively intact age- and education-matched control participants (NC) produced spoken word-associations in response to single spoken word stimuli. In both experiments participants with AD produced responses that were less common than NC participants' responses. For example, given the stimulus *bride* nearly half (46%) of participants in the NC group, but only 25% of the participants in the AD group, said *groom* (the most strongly associated response according the Nelson et al., 1998, norms). However, only stimulus words that had a very strong associate in the norms, that is, strong stimuli, produced a difference between groups (an example of a strong stimulus is *bride* which led to the response *groom* in 87% of the Nelson et al., normative sample). In contrast, weak stimuli, that is, words that did not have a single strongly associated response in the norms, produced no difference between groups. For instance, given the stimulus *confusion* no participant (in either group) produced *lost* (which 7% of the normative sample produced and which was the strongest association to *confusion* in the norms), and the same number ($n = 5$ or 28%) of NC and AD participants produced one of the next the strongest associates (i.e., *understand, mix, dismay, or mess* all of which 3% of the normative sample produced).

These results confirmed the prediction of the degraded semantics hypothesis; participants with AD produced different types of associations relative to controls. However, the hypothesis does not necessarily predict that stronger associations should more readily demonstrate the effects of cognitive impairment. In fact, assuming that strong associations are more robust, this finding is somewhat unexpected because it could be argued that stronger associations should be more resistant to damage (e.g., Balota, Watson, Duchek, & Ferraro, 1999). Before considering how to explain this result, it is important to establish that association strength was the critical factor.

Several aspects of the data suggest that association strength determined whether or not patients with AD were impaired on the word-association task rather than a between group difference in approach to the task, or some property of the stimuli that was confounded with stimulus strength. In Experiment 2, participants produced more than one association to the same stimulus word thereby effectively converting stimuli with strong associates into weak stimuli (because only weak associates remain after the strong associate is produced). The results showed that even responses to strong stimuli only produced a difference between AD and NC participants on the first association; using the very same strong stimuli, no differences emerged between participant groups on the second and third association trials. An analysis of 40 randomly selected data points from Experiment 2 also provided no evidence that participants were chaining their responses on the second and third association trials. Instead, it seems that the AD and NC participants produced equally common responses to strong stimuli after the strongest association was exhausted (by the first response). These data strongly suggest that stimulus strength was critical, and imply that strong associations are more vulnerable to the cognitive effects associated with AD.

Other analyses provided evidence against the possibility that participants with AD produced less common responses (to strong stimuli on the first response) simply because they failed to understand, or could not perform, the word association task. Subjective ratings of the relationship between each stimulus and each response demonstrated that even when participants with AD did produce responses that were (judged to be) semantically related to the stimulus, these were relatively less common than responses that the NC participants produced. In other words, to the extent that participants with AD performed the task as instructed, their responses were nevertheless qualitatively different (again only in the strong stimulus condition).

A few additional analyses provide evidence against another, relatively uninteresting, possible reason why weak associations were not sensitive to AD, which is that weak associations were subject to a floor effect, or that weak stimuli lacked the power to reveal group differences of any kind. First, in Experiment 2 medium strength associations were not at floor and also produced no effect of AD (see Section 3.2). It could be argued that even medium strength stimuli produced responses that were close to the floor, however, the magnitude of the participant group difference in response typicality on strong stimuli in Experiment 1 was only 8% (41 minus 33; see Table 2), and as such there was room in responses to the middle strength stimuli to produce a similarly sized effect.

Second, although weak stimuli did not demonstrate sensitivity to the cognitive changes associated with AD, both strong and weak associations were sensitive to age-differences[3] particularly in in Experiment 1 in which the age range was 60–89. Individuals closer to 90 produced less common responses when given strong and weak stimuli. The correlation between participant age and response commonality was $r^2 = -.249$, $p < .01$ for strong and $r^2 = -.192$, $p = .01$ for weak stimuli. The same analysis of the data from Experiment 2 (using responses on the first association trial only) showed no significant correlation between age and response commonality of $r^2 = .07$, $p = .194$; however, relative to Experiment 1 there were both fewer subjects and fewer items in Experiment 2—a difference which may have interfered with the possibility of detecting an age effect. Where the age effect was obtained (i.e., in Experiment 1) age was not significantly correlated with education ($r^2 = .02$, $p = .436$) and also not with raw scores on the WAIS-R Vocabulary subtest ($r^2 = 0.004$, $p = .737$; for this correlation no vocabulary score was available for 3 of the NC participants) suggesting that the age effect was not simply an artifact of age differences in vocabulary (as in previous comparisons of older to younger adults; Burke & Peters, 1986). In contrast, strong stimuli in Experiment 1 produced a correlation with DRS scores $r^2 = 0.234$, $p < .01$ whereas weak stimuli did not $r^2 = .016$, $p = .46$ again indicating that strong but not weak stimuli were sensitive to the cognitive changes associated with AD. The age related reduction in response commonality could reflect retrieval problems or the fact that increased age entails different experiences and therefore also differences in semantic memory (see below). These data demonstrated that weak stimuli could detect some types of group differences.

[3] Note that the participants in the current study were much older than the undergraduates who produced the Nelson norms. One way to test if strong and weak stimuli were sensitive to age effects would be to compare responses from the current study to those in the Nelson norms. However, given that the participants in the current study would likely differ from those who generated the Nelson norms in a number of other variables we refrained from carrying out such analyses.

A third piece of evidence that association strength, and not a floor effect, was critical was that within the strong items (the stimuli that demonstrated robust differences between AD participants and controls) there was a relationship between association strength and the group difference. This analysis is shown in Fig. 1. Stronger associations produced bigger group differences, and relatively few stimuli produced effects in the wrong direction (more typical responses in participants with AD which in Fig. 1 are difference scores below zero).

4.1. The degraded-semantics hypothesis

The results we obtained confirmed the general prediction that semantic associations should be different in participants with AD than in normal controls. However, further explanation is needed to explain why the difference was restricted to strong stimuli, and why participants with AD produced less typical (instead of more typical) responses than controls. One view of the association task would be consistent with the degraded semantics hypothesis and would also explain the interaction between participant type and stimulus strength is suggested by the subjective analysis which showed that weak stimuli were significantly less likely to produce semantically related responses (in both the AD and NC groups).

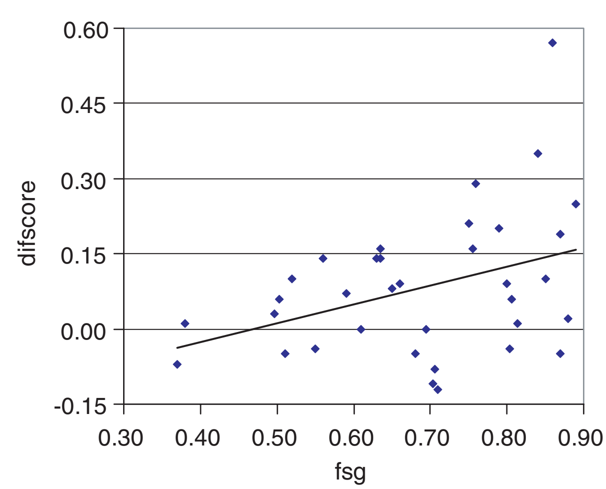

Fig. 1. The difference in response typicality (difscore on the *Y* axis) between the AD and NC groups (AD minus NC typicality scores) plotted as a function of a continuous measure of association strength (the variable FSG on the *X* axis) using first responses to the thirty-five stimuli with the strongest associates in Experiments 1 and 2. Within the strong items, words with stronger associates produced a significantly greater difference between participant groups ($r^2 = 0.14$). The relationship between participant type and association strength as a continuous variable within the strong items provides further evidence that strong associations are more sensitive to the effects of AD than weaker associations (and suggests that the interaction between association strength and participant type is not merely an artifact of a floor effect in the weak items; see Tables 2 and 4 and Fig. 2).

Fig. 2 shows how the results differed when we classified the responses subjectively (i.e., categorically as either semantically related or not) versus when we used typicality ratings. For this figure we combined the data from all the participants in Experiments 1 and 2 (leaving out the data from 2nd and 3rd association trials and the data on items with medium strength associates in Experiment 2). A small number of subjects participated in both Experiments 1 and 2 (4 with AD and 3 from the NC group). These participants' data were entered only once using the materials from Experiment 1 (which had many more stimulus items than Experiment 2).

The top panel of Fig. 2 shows the interaction; strong stimuli produced a difference in response typicality (using the Nelson et al., 1998 norms) between AD and NC participants, but weak stimuli produced no difference between groups. In contrast, the bottom panel of Fig. 2 shows that, for both participant types, responses to weak stimuli were less likely to be judged as semantically related to the stimulus than responses to strong stimuli.

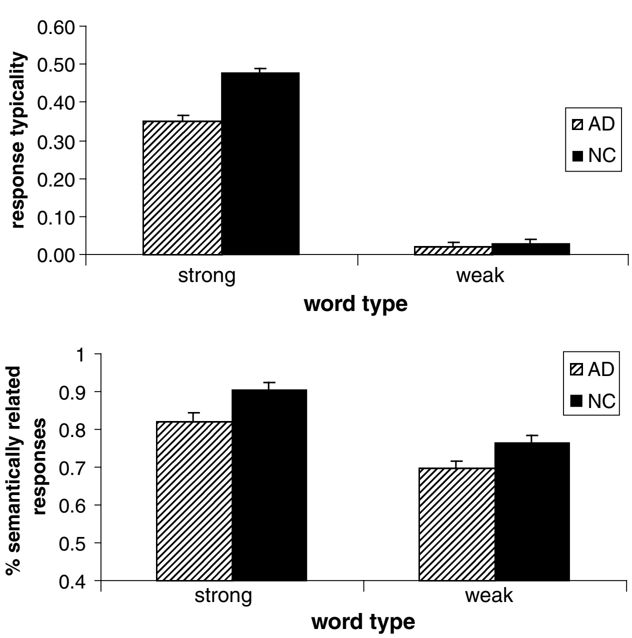

Fig. 2. In the top panel, mean typicality of responses to words with strong versus weak associates (typicality based on the variable FSG in the Nelson et al., 1998, norms). In the bottom panel, mean proportion of responses that were semantically related to the stimulus (based on subjective judgments). (Error bars show standard errors for the participant group means; data in both panels were collapsed across Experiments 1 and 2.) Participants with AD produced less typical responses to words with strong associates (top panel), and fewer semantically related responses overall (bottom panel). In both the AD and the NC group, responses to words with weak associates were less likely to be semantically related to the stimulus word (bottom panel). Together these findings suggest that words that are weakly associated with many other words activate concepts less than words that are strongly associated with just one word.

The subjective analysis (shown in the bottom panel of Fig. 2) suggests that different cognitive processes may be involved in producing word associations to strong versus weak stimuli. Specifically, participants may be less likely to rely on meaning when producing associations to weak stimuli. By this view, even when semantically related responses were produced these involved meaning to a lesser extent for weak than for strong stimuli. If weak associations are effectively "less semantic" in nature, and AD primarily entails damage to semantic representations, then the AD effect might be restricted to strong stimuli. This account of the word-association task accounts for the counter-intuitive aspect of the findings (that stronger associations were more affected by cognitive impairment) by positing that strong associations entail more of the type of processing that is affected by AD. This view of the word association task is also consistent with previous suggestions that "strong" representations (e.g., high frequency words) will be less vulnerable to structural degradation (e.g., Hutchison & Balota, 2003). The reason why weak stimuli did not demonstrate an effect of AD is not because weak stimuli were structurally intact, but because intact semantic representations may not be as important for producing responses to weak stimuli as they may be for producing responses to strong stimuli.

A necessary next step is to specify what it means to say that strong stimuli involved increased semantic processing. The results of Experiment 2 suggested that "more semantic" in this case is unlikely to reflect something as simple as "more concrete" because the same words (or concepts) demonstrated the effect on the first but not on second and third association responses (i.e., not after strong associations were exhausted). One possibility is all or none: weak stimuli may be more likely to activate associations that are not based on meaning at all. However, the bottom panel of Fig. 2 demonstrates that the majority of responses, even to words with only weak associations, were semantically related to the stimulus. As such, the proposal that strong associations entail more semantic processing than weak associations may need to apply in a continuous rather than an all-or-none fashion. That is, even when weak stimuli do lead to semantically related responses these may entail relatively less activation of meaning than strong associations. As an example, this could be operationalized in some distributed models of meaning representation (e.g., Cree, McRae, & McNorgan, 1999; Masson, 1995; McRae, de Sa, & Seidenberg, 1997) by having increased numbers of overlapping semantic features for strong stimuli and their associations relative to weak stimuli and their associations. Within this type of framework, participants with AD might be expected to produce less typical responses if small amounts of damage had a greater effect on strong associations effectively reducing them in strength such that they no longer differed from weaker associations.

That the majority of responses were semantically related to the stimulus is also consistent with our assumption that meaning drives responses in the word-association task. Without this assumption the degraded semantics hypothesis would predict that participants with AD should produce responses similar to NC participants in the word-association task. An alternative possibility is that lexical associations between word forms play a role in the word association task. In many cases it would be impossible to distinguish between the semantic and lexical-associative alternatives because word pairs can be associated both semantically and at the level of word forms. In some cases, lexical associations seem obviously important. For example, the strongest association to the weak stimulus *chicken* was *soup*. The semantic properties of *soup* are hardly related to those of *chicken* but the words "*chicken soup*" do occur together.

We cannot rule out the possibility that associations between word forms influence responses, at least in part (this is a common problem in psycholinguistic research; Hutchison, 2003; Thompson-Schill, Kurtz, & Gabrieli, 1998). Previous investigators, however, argued that meaning drives responses in the word association task (de Groot, 1989), and as already noted (see Section 1) consistent with this proposal bilinguals demonstrated similar associations between languages even though words in different languages seldom occur together (van Hell & de Groot, 1998). In addition, if associations between word forms were involved, then it is odd that so few (less than or approximately 1%; see Table 3) of responses in the association task were purely form related to the stimulus (e.g., *cat* and *can*; see also Postman & Keppel, 1970).

4.2. The retrieval account

Some investigators suggest that semantic impairments in AD reflect a processing deficit that interferes with information retrieval from a relatively intact semantic memory store (for reviews, see McGlinchey-Berroth & Milberg, 1993; Nebes, 1989; Nebes, 1992; Ober, Shenaut, & Reed, 1995). According to this view, the structure of semantic memory remains essentially unchanged in the early stages of the disease, and the deficits exhibited on standard language tasks (e.g., confrontation naming, verbal fluency) result from difficulty with retrieving the desired lexical representations.

The retrieval-based account predicts that AD related deficits should be particularly obvious when retrieval is difficult. In this respect, it is difficult for the retrieval account to explain why strong associations (which seem to immediately and easily pop to mind) were sensitive to AD whereas weak associations were not. This reversed pattern (relative to the current set of results) was obtained in one study; strong but not weak associations produced intact priming in participants with AD (Balota et al., 1999). The retrieval account cannot side step the current results (as above) by proposing that a property of the word-association task caused strong but not weak stimuli to demonstrate an effect of AD; retrieval is clearly equally necessary to produce a response to a weak and a strong stimulus. The retrieval account is based primarily upon the results of studies that demonstrated intact semantic priming in patients with AD (Nebes, Martin, & Horn, 1984; Ober, Shenaut, Jagust, & Stillman, 1991). To some extent the difference in findings across paradigms may result from the fact that different processing mechanisms are needed in different tasks.

Table 6
The mean response frequency per million produced by participants in each group in Experiments 1 and first-response trial in Experiment 2

Group type	Experiment 1		Experiment 2	
	Stimulus type		Stimulus type	
	Strong ($n = 26$)	Weak ($n = 26$)	Strong ($n = 6$)	Weak ($n = 6$)
AD	151 (80)	174 (139)	248 (165)	98 (52)
NC	105 (46)	122 (41)	147 (77)	103 (57)
Difference	46.85	51.82	100.84	−4.5
Two-tailed t test[a]	2.16	1.51	1.99	<1
p value	0.04	0.14	0.06	0.22

Standard deviations are shown in parentheses.
[a] Degrees of freedom = 34 in Experiment 1 and 23 in Experiment 2; data presented are analyzed over subjects not items.

To assess the extent to which retrieval difficulty contributed to the current findings, we used word frequency as a marker of retrieval difficulty and conducted an analysis to determine if participants with AD produced higher frequency words relative to those in the NC group. If participants with AD had more difficulty retrieving the most common associate they may have produced higher frequency (but less common) associates instead. The results of this analysis are shown in Table 6. Consistent with this proposal, participants with AD produced higher frequency words relative to the NC participants. However, this effect was significant only for strong stimuli again suggesting that strong associations were more vulnerable to the effects of AD relative to weak associations—a finding that is the opposite of what the retrieval difficulty account predicts. The observed pattern could not be attributed to differences in baseline frequency because strong ($M = 120.91$, $SD = 128.65$) and weak ($M = 89.40$, $SD = 110.89$) stimuli did not differ $t(64) = 1.05$, $p = .30$ on the average frequency of the strongest associate in the Nelson et al. (1998) norms (for this analysis we combined the 56 stimuli used in Experiment 1 and the 6 strongest and 6 weakest stimuli used in Experiment 2).

An additional variable to consider in the context of the retrieval deficit hypothesis is the number of different associates. One recent study elaborated the retrieval account with evidence that participants with AD have a specific difficulty selecting one response from a number of possible responses (i.e., a selection deficit). Impressively, in one experiment the AD effect was greater on a relatively easier task (fluency for the letter F) and the AD related deficit was smaller on a more difficult task (fluency for words that begin with FL). Because there are more F words than FL words, and therefore more response competition in generating F words (than FL words), this result suggests that participants with AD have particular difficulty when competition for selection is greater (Tippett, Gendall, Farah, & Thompson-Schill, 2004).

Assuming that normative set size is an indication of how many competitors become active in each individual cognitive system during the word association task (for discussion of this point see Nelson et al., 2000), then the selection deficit view of AD would also predict that weak stimuli should have produced a bigger (not smaller) participant difference in the current study because strong stimuli had fewer associates (both in the current study and in the Nelson et al., 1998, norms). However, we did not manipulate or control the number of associations and this leaves open the possibility that both association strength (FSG) and association set size (QSS) may affect responses in this task (the correlation between FSG and QSS for all the materials in the current study was $r^2 = -.90$). To determine if set size has an effect it would be necessary to manipulate set size while controlling association strength.

The age related reduction in response commonality to both strong and weak stimuli is also potentially useful for distinguishing between the retrieval and semantic deficit hypotheses. Research on aging converges on the conclusion that semantic representations are intact in older adults whereas advanced age impairs the ability to access word forms for production (for a review see Burke, MacKay, & James, 2000). Given that increased age in the current study affected responses to both strong and weak stimuli it would seem that both types of stimuli are sensitive to retrieval deficits where they exist. As such, both types of stimuli should have been sensitive to AD if AD primarily entails problems with retrieval. This conclusion, however, assumes that the observed age effects on word association on both strong and weak stimuli resulted from increased difficulty with retrieving the word forms. An equally plausible explanation for the age related reduction in response commonality is that word associations differ as semantic representations continue to develop over the life-span (Hirsh & Tree, 2001). In other words, the age effect is not unambiguously related to retrieval deficits, or deficits of any kind.

Another approach to distinguishing between the degraded semantics and retrieval-based accounts is to examine word-associations in patients with identified retrieval deficits. Huntington's disease is a dementing neurodegenerative disorder considered by some (Butters et al., 1987; Butters, Wolfe, Martone, Granholm, & Cermak, 1985; Monsch et al., 1994), to be characterized by a general retrieval deficit (but see Barr & Brandt, 1996; Suhr & Jones, 1998). Notably, Randolph (1991) reported that relative to cognitively intact controls, patients with Huntington's disease generated equally common word associations. If the above-reported sensitivity of strong associations to AD resulted from a retrieval deficit, then patients with Huntington's disease should have demonstrated a similar pattern of performance. This suggests that the word-association task is not sensitive to retrieval deficits, and as such the AD related reduction in response commonality should instead be attributed to changes in the structure of semantic memory, and processing differences involved in generating associations to weak versus strong stimuli.

4.3. Clinical implications

To determine if the word association task may be useful in clinical settings we conducted a logistic regression with deviation from the NC mean as the independent variable and classification into the AD versus the NC groups as the dependent variable. For this analysis we used only the responses to strong stimuli (because these were sensitive to the effects of AD), and we combined the data from Experiments 1 and 2 (again excluding the Experiment 2 data for 7 subjects who participated in both experiments). We carried out the logistic regression twice, once using performance that was a single standard deviation below the NC mean as the independent variable, and once using two standard deviations. Both analyses produced significant findings but we report only the single standard deviation results because these were slightly more accurate in predicting the probability of AD. The Chi-squared for the total model was adequate ($p = .005$) and the total correctly classified was 68.52% (or 82% of the NC and 54% of AD participants). The inverse of the estimated odds-ratio (1/Exp*B*) indicated that participants who produced words that were on average 1 *SD* below the NC mean (associative strength according to the Nelson et al., 1998 norms) were 5.38 times (CI 1.56–18.49) more likely to have AD than those who produced associates at or above the NC mean ($Wald = 7.09$, $p = 0.001$). This suggests that the association test may be clinically useful for distinguishing between people with AD and normal controls but the results would first need to be replicated in larger groups of participants.

5. Conclusions

The vulnerability of strong but not weak associations to AD, the reduced typicality of responses in AD relative to NC participants, and the fact that weak stimuli led to fewer semantically related responses relative to strong stimuli (in both the AD and NC groups) constrain hypotheses about the nature of cognitive impairments in AD as well as theories of how meaning is represented within the normal cognitive system. The current data do not support retrieval-based accounts of the deficits in AD because strong associates are easier to access and produce, and these should thus have been less (not more) vulnerable to retrieval deficits. The current data also constrain accounts of the word association task. To explain why strongly associated concepts should be more vulnerable to AD and changes in semantic representations relative to weakly associated concepts we suggested that weak associations are less dependent on meaning than strong associations. This proposal predicts that early AD will be more likely to affect tasks that entail more semantic processing, and that weak stimuli entail less semantic processing than strong stimuli even in cognitively intact people. On a more general note, the current study demonstrated how data from cognitively impaired populations can lead to potentially useful clinical tests while also improving our understanding of normal cognitive processes.

Appendix

Experiment 1

Strong stimuli	Weak stimuli
ashtray	body
astronomy	chicken
bait	cleaner
broth	computer
chlorine	conceit
circle	condemn
cork	confusion
crib	crisis
dill	culture
duplicate	disown
error	extreme
flipper	farmer
husband	field
keg	fray
lens	grace
option	intensity
sap	mastery
shingle	mutton
skunk	natural
slippery	obscure
stumble	overwhelm
throne	range
tribe	renounce
weapon	resistance
whiskers	standard
yoke	tact

Experiment 2

Strong stimuli	Middle stimuli	Weak stimuli
bride	bride	
cash	direction	
father	eggs	lizard
library	harvest	relax
margarine	minutes	secret
scissors	winter	wall

References

Abeysinghe, S. C., Bayles, K. A., & Trosset, M. W. (1990). Semantic memory deterioration in Alzheimer's subjects: Evidence from word association, definition, and associate ranking tasks. *Journal of Speech and Hearing Research, 33*, 574–582.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. University of Pennsylvania: Philadelphia Linguistic Data Consortium.

Balota, D. A., Watson, J. M., Duchek, J. M., & Ferraro, F. R. (1999). Cross-modal semantic and homograph priming in healthy young, healthy old, and in Alzheimer's disease individuals. *Journal of the International Neuropsychological Society, 5*, 626–640.

Barr, A., & Brandt, J. (1996). Word-list generation deficits in dementia. *Journal of Clinical and Experimental Neuropsychology, 18*, 810–822.

Burke, D. M., MacKay, D. G., & James, L. E. (2000). Theoretical approaches to language and aging. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 204–237). Oxford, UK: Oxford University Press.

Burke, D. M., & Peters, L. (1986). Word associations in old age: Evidence for consistency in semantic encoding during adulthood. *Psychology and Aging, 1*, 283–291.

Butters, N., Granholm, E., Salmon, D. P., Grant, I., & Wolfe, J. (1987). Episodic and semantic memory: A comparison of amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology, 9*, 479–497.

Butters, N., Wolfe, J., Martone, M., Granholm, E., & Cermak, L. (1985). Memory disorders associated with Huntington's disease: Verbal recall, verbal recognition, and procedural memory. *Neuropsychologia, 23*, 729–743.

Chertkow, H., & Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type. *Brain, 113*, 397–417.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science, 23*, 371–414.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 824–845.

Eustache, F., Cox, C., Brandt, J., & Lechevalier, B. (1990). Word-association responses and severity of dementia in Alzheimer disease. *Psychological Reports, 66*, 1315–1322.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Gewirth, L., Shindler, A., & Hier, D. (1984). Altered patterns of word associations in dementia and aphasa. *Brain and Language, 21*, 307–317.

Hachinski, V. C., Iliff, L. D., Zilhka, E., DuBoulay, G. J., McAllister, V. L., Marshall, J., et al. (1975). Cerebral blood flow in dementia. *Archives of Neurology, 32*, 632–637.

Hirsh, K., & Tree, J. J. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics, 14*, 1–44.

Hutchison, K. A. (2003). Is semantic priming due to association strength on featural overlap? A micro-analytic review. *Psychonomic Bulletin & Review, 10*, 785–813.

Hutchison, K. A., & Balota, D. A. (2003). Structure versus processing deficits in Alzheimer's disease, a matter of degree: A comment on Storms et al., 2003. *Neuropsychology, 17*, 306–309.

Martin, A. (1992). Degraded knowledge representation in patients with Alzheimer's disease. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (2nd ed., pp. 220–232). New York: Guilford Press.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 3–23.

Mattis, S. (1988). *Dementia Rating Scale: Professional Manual*. Odessa, Florida: Psychological Assessment Resources.

McGlinchey-Berroth, R., & Milberg, W. P. (1993). Preserved semantic memory structure in Alzheimer's disease. In J. Cerella, J. Rybash, W. Hoyer, & M. L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 407–422). New York: Academic Press.

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. *Neurology, 34*, 939–944.

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

Monsch, A. U., Bondi, M. W., Paulsen, J. S., Brugger, P., Butters, N., Salmon, D. P., et al. (1994). A comparison of category and letter fluency in Alzheimer's disease and Huntington's disease. *Neuropsychology, 8*, 25–30.

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

Nebes, R. D. (1992). Cognitive dysfunction in Alzheimer's disease. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 373–446). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nebes, R., Martin, D., & Horn, L. (1984). Sparing of semantic memory in Alzheimer's disease. *Journal of Abnormal Psychology, 93*, 321–330.

Nelson, D. L., McEvoy, C. L., & Dennis, S. (2000). What is free association and what does it measure?. *Memory & Cognition, 28*, 887–899.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. [On-line]. http://cyber.acomp.usf.edu/FreeAssociation/.

Ober, B. A., Shenaut, G., Jagust, W. J., & Stillman, R. C. (1991). Automatic semantic priming with various category relations in Alzheimer's disease and normal aging. *Psychology and Aging, 6*, 647–660.

Ober, B. A., Shenaut, G. K., & Reed, B. R. (1995). Assessment of associative relations in Alzheimer's disease: Evidence for preservation of semantic memory. *Aging and Cognition, 2*, 254–267.

Postman, L., & Keppel, G. (1970). *Norms of word association*. New York: Academic Press.

Randolph, C. (1991). Implicit, explicit, and semantic memory function in Alzheimer's disease and Huntington's disease. *Journal of Clinical and Experimental Neuropsychology, 13*, 479–494.

Randolph, C., Braun, A. R., Goldberg, T. E., & Chase, T. N. (1993). Semantic fluency in Alzheimer's, Parkinson's, and Huntington's disease: Dissociation of storage and retrieval failures. *Neuropsychology, 7*, 82–88.

Rohrer, D., Wixted, J. T., Salmon, D. P., & Butters, N. (1995). Retrieval from semantic memory and its implications for Alzheimer's disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1127–1139.

Rosen, W. G., Terry, R. D., Fuld, P. A., Katzman, R., & Peck, A. (1980). Pathological verification of ischemia score in differentiation of dementias. *Annals of Neurology, 7*, 486–488.

Salmon, D. P., Butters, N., & Chan, A. S. (1999). The deterioration of semantic memory in Alzheimer's disease. *Canadian Journal of Experimental Psychology, 53*, 108–117.

Salmon, D. P., Heindel, W. C., & Lange, K. L. (1999). Differential decline in word generation from phonemic and semantic categories during the course of Alzheimer's disease: Implications for the integrity of semantic memory. *Journal of the International Neuropsychological Society, 5*, 692–703.

Santo Pietro, M. J., & Goldfarb, R. (1985). Characteristic patterns of word association responses in institutionalized elderly with and without senile dementia. *Brain and Language, 26*, 230–243.

Suhr, J. A., & Jones, R. D. (1998). Letter and semantic fluency in Alzheimer's, Huntington's, and Parkinson's dementias. *Archives of Clinical Neuropsychology, 13*, 447–454.

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. E. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language, 38*, 440–458.

Tippett, L. J., Gendall, A., Farah, M. J., & Thompson-Schill, S. L. (2004). Selection ability in Alzheimer's disease: Investigation of a component of semantic processing. *Neuropsychology, 18*, 163–173.

van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language ; Cognition, 1*, 193–211.