Early assessment of dementia: the contribution of different memory components

Spaan, P.E.J.

Citation for published version (APA):
Spaan, P. E. J. (2002). Early assessment of dementia: the contribution of different memory components

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER V: NATURE OF PERFORMANCE IN THE ‘CATEGORY FLUENCY TEST’: AN EXPLORATION OF THE SPECIFIC RESPONSES

1. Introduction

Verbal, or category, fluency is a task frequently applied in clinical neuropsychological assessment as well as experimental neuropsychological research, in which the patient (or subject) is asked to name as many exemplars of a presented category within a certain time limit (usually one minute). In clinical assessment, this task is used as a measure of several cognitive functions: sometimes it is used as a language test (i.e., verbal expression; Lezak, 1995), while it is also frequently considered to be an executive functioning test (e.g., Hanes, Andrewes, Smith & Pantelis, 1996). Experimental research primarily considers this task to be a measure of semantic memory (i.e., semantic retrieval). It may be noted that category fluency is a complex task dependent upon several cognitive processes including attention and retrieval strategies, working memory, intact semantic stores and phonological processes (e.g., Chertkow & Bub, 1990).

The complexity of the category fluency task is also reflected by neuro-imaging research. For example, Frith, Friston, Liddle and Frackowiak (1991) found that in four normal subjects intrinsic generation of words (i.e., verbal fluency) was associated with an increase in left dorsolateral prefrontal cortical activity and a bilateral decrease in activity in auditory and superior temporal cortices. In addition, Troyer (2000) reviews that optimal verbal fluency performance mainly involves two components of cognitive functioning: ‘clustering’ (i.e., generating words within a subcategory – a process dependant on semantic categorisation) and ‘switching’ (i.e., cognitive flexibility in shifting from one subcategory to another). Clustering may be related to temporal lobe functioning, whereas switching is related to frontal lobe functioning.

As was reviewed in Chapter I, AD patients are found to perform significantly worse in category fluency tests than normal elderly controls (e.g., Hodges, Salmon & Butters, 1990; Monsch et al., 1992; Mickanin, Grossman, Onishi, Auriacombe & Clark, 1994; Hodges & Patterson, 1995; Beatty, Testa, English & Winn, 1997; Sailor, Bramwell & Griesing, 1998). In addition to naming few correct exemplars in general, AD patients typically name the most common elements and produce few different subcategory clusters, few items per subcategory and relatively many category labels (e.g., Martin & Fedio, 1983; Ober, Dronkers, Koss, Delis & Friedland, 1986; Tröster, Salmon, McCullough & Butters, 1989; Beatty et al., 1997).

Chertkow and Bub (1990) found that category fluency was impaired in their AD subjects due to two major constraints: deterioration of semantic memory store and difficulties in semantic search. Semantic search problems (i.e., impaired access) might be affected by attentional problems, motivation and the ability to devise strategies for breaking down categories into smaller sections for efficient search. In addition, AD patients presumably suffer from a specific deterioration (or degradation) of semantic knowledge stores, characterised by a difficulty in differentiating between items within the same semantic category concurrent with relative preservation of broader categorical information (Martin & Fedio, 1983). Thus, AD results in a loss of knowledge concerning those attributes that distinguish objects within the same
Nature of performance in the ‘Category fluency test’...

semantic category, while they can make accurate judgments concerning superordinate and specific category membership (Chertkow & Bub, 1990). This is called the bottom-up breakdown of semantic knowledge and is often used to explain the nature of performance of AD patients (e.g., Ober et al., 1986; Salmon, Shimamura, Butters & Smith, 1988; Tröster et al., 1989; Monsch et al., 1994; Rosser & Hodges, 1994; Binetti et al., 1995). AD patients’ semantic representation of concepts is assumed to be abnormally underspecified because of a lack of critical, object-specific attributes (Martin, in Squire & Butters, 1992). The hypothesis of bottom-up semantic breakdown corresponds with the hierarchical organisation of semantic knowledge, proceeding from knowledge of specific attributes to more global, superordinate information (Warrington, 1975).

However, most studies investigating category fluency performance in AD patients involve patients whose diagnosis has been made with rather great certainty, though they were still in an early stage of the disease. The current research concerns the investigation of memory performance of preclinical AD patients, relative to cognitively healthy elderly subjects. Few studies focus on the efficiency of category fluency tasks in early assessment of dementia. It is, therefore, not clear how early in the progression of the disease the deficits in category fluency can be detected. Nonetheless, Weingartner, Kawas, Rawlings and Shapiro (1993) found that changes in semantic memory were detectable before the diagnosis of AD could be made: normal elderly controls generated more uncommon exemplars from semantic categories (fruits and vegetables) than AD patients did 2½ years prior to the presumed onset of AD. First, the patients were not able to name middle- and low-frequency exemplars and later – when the diagnosis of AD was made – they lost more common, high-frequency, elements (frequency norms derived from Battig & Montague, 1969). The subjects were derived from a longitudinal ageing study, as was done in the current research, but only 6 AD patients and 6 matched normal controls were available for analysis of semantic memory and fluency performance.

The current research takes an effort to investigate the nature of performance on the ‘Category fluency test’ of the heterogeneous sample of elderly subjects. First, the two components that determine successful performance on a category fluency paradigm – clustering and switching (e.g., Troyer, 2000) – will be analysed. The clustering measure reflects the ability to generate words within a subcategory, while the switching measure refers to the ability to shift to another subcategory when one cannot retrieve more words from the previous subcategory. Deficits in these two components may represent, in fact, semantic search problems.

Secondly, the degree of integrity of semantic store will be investigated by examining the frequency distribution of responses generated by the subjects. It will be analysed whether the subjects generate relatively few uncommon (middle- and low-frequency) exemplars of a category, compared with common (high-frequency) exemplars. This allows for an investigation of the hierarchical organisation of semantic knowledge of the elderly subjects, which might indicate a bottom-up breakdown, as is frequently found in AD.

These aspects of category fluency performance will, first, be investigated in the cognitively impaired (CI) subjects and their matched controls (mNC), known from the first administration of the ‘Category fluency test’ (T1) (section 3). Thus, this section will determine the influence of global level of cognitive functioning (measured by the MMSE) on various aspects of category fluency performance within a heterogeneous group of clinically non-demented elderly subjects. Secondly, the
development of the various scores over time will be discussed when analysing the data derived from the second administration, two years later (T2-T1) (section 4). It will be investigated whether the various scores show decline or a stable pattern of performance once subjects are diagnosed as demented at T2, relative to the non-demented subjects. In addition, in section 5, the predictive value of the various scores with respect to dementia will be discussed (i.e., sum of correct responses, number and nature of incorrect responses, scores indicating semantic search problems, and scores indicating the integrity of semantic store). An attempt will be made to replicate the findings by Weingartner et al. (1993): do subjects diagnosed as demented at T2 generate fewer uncommon exemplars two years before diagnosis than subjects that turned out to be non-demented? It will be examined whether the extra scores derived from the ‘Category fluency test’ improve the prediction of the demented cases, relative to the traditional score of this test, the sum of correct exemplars. Furthermore, the contribution of these scores in the prediction of dementia, relative to the other measures of the memory test battery (see Chapter IV), will be determined.

2. Method

The global procedure and scoring method of the ‘Category fluency test’ is described in Chapter II. In addition to determining the correct (unique) responses (i.e., responses that are no intrusions, repetitions or ‘same exemplar’ perseverations), each correct response was classified to the most relevant subcategory type, in order to examine the clustering and switching nature of performance of the elderly subjects (section 2.1). Furthermore, each response was characterised in terms of its frequency of occurrence, in order to investigate the frequency distribution of the answers (section 2.2).

2.1. Switching and clustering scores

First, it should be noted that the classification of the animals and occupations to different subcategories is still in an experimental phase. Extensive research using a category fluency paradigm in AD patients and normal elderly subjects focuses on the category ‘supermarket items’, rather than on ‘animals’ or ‘occupations’ (e.g., Martin & Fedio, 1983; Ober et al., 1986; Tröster et al., 1989). Indeed, classification of answers to different subcategories within the category of ‘supermarket items’ is less ambiguous, since in a real supermarket the items are also classified to different departments (i.e., subcategories). No studies were found that used the category ‘occupations’ in AD patients and only a few studies had classified answers to the category ‘animals’ in different subcategories – though the exact subcategories differed per study (Randolph, Braun, Goldberg & Chase, 1993; Binetti et al., 1995; Beatty et al., 1997; Troyer, 2000). Nevertheless, the limitation to ‘supermarket items’ is regrettable since in clinical neuropsychological assessment, use of the category fluency paradigm usually involves ‘animals’ and ‘occupations’ (except for the Dementia Rating Scale (M Mattis, 1988) which involves ‘supermarket items’). However, clinical assessment focuses on determining the number of correct responses rather than additional scores that may provide information on the specific response profile of a patient.

Despite all these complications, an attempt was made to categorise the responses generated in the animal and the occupation category (see Appendix A and B for the division of different subcategories of the animals and the occupations,
respectively). In this way, it was possible to determine the number of different subcategories that the subject intrinsically produced, which served as a measure for switching. In addition, the sum of correct responses was divided by the score indicating the number of different subcategories: (total number of correct responses) / (number of different subcategories). This proportion indicated the average size of a ‘cluster’ of related exemplars and, thus, served as a measure for clustering.

2.2. Frequency distribution of responses

In order to investigate the responses generated by the subjects in terms of their frequency distribution, each response was scored as a high-frequency (HF) answer, a middle-frequency (MF) answer or a low-frequency (LF) answer. Only two studies were found that administered a category fluency task and explicitly examined the frequency distribution of the responses generated in reaction to the category presented. The first study, by Binetti et al. (1995), simply examined the general lexical frequency of the exemplars generated (i.e., mean normalised frequency of usage according to a lexicon of spoken Italian), apart from frequency of occurrence within the specific category (i.e., animals). The second study, Weingartner, Kawas, Rawlings & Shapiro (1993), used the categories of fruits and vegetables and based the frequency of responses on frequency norms provided by Battig and Montague (1969). However, the research by Battig and Montague represents performance of a population of young college students, which differs essentially from the nature of the current research sample (elderly subjects with typically primary school education). In addition, no norms were available for the category ‘animals’ – Battig and Montague ask for ‘four-footed animals’, ‘birds’, ‘insects’ and ‘fish’ in separate trials instead of joined together in one trial.

Therefore, because of the lack of available appropriate norms, the responses generated by our own sample of elderly subjects were used to determine the HF, MF and LF responses. This sample was limited to the normal control subjects (the mNC and the nNC group; n=82) in order to prevent influence of the possible pathological nature of performance of the CI subjects. LF responses can be characterised as answers generated only once or twice by the sample of 82 elderly control subjects, both in the animal category and in the occupation category. MF responses in the animal category are answers generated 3-17 times by the sample of 82 elderly controls; MF responses in the occupation category are answers generated 3-9 times. HF responses in the animal category are answers generated 19 times or more (maximum of 64 times in the sample of 82 elderly controls); HF responses in the occupation category are answers generated 10 times or more (maximum of 45 times). Different absolute frequency limits were applied for the animal and the occupation category, since the respective frequency distributions and response profiles differed – i.e., more answers in general were generated in animal category, while more different answers were generated in the occupation category. The general rule applied was that the HF responses should reflect approximately 50% of the responses generated by a subject. Appendix A and B offer an impression of the frequency of occurrence of the specific responses in both categories.
3. Results and discussion of the first administration of the ‘Category fluency test’ – a comparison of scores between the cognitively impaired and the normal control subjects

In this section, the various scores derived from the ‘Category fluency test’ will be compared between the CI and the mNC subjects. Thus, the influence of global level of cognitive functioning (measured by the MMSE) on various aspects of category fluency performance within a heterogeneous group of clinically non-demented elderly subjects will be determined. For details on the demographic or cognitive status features of the CI and the mNC subjects, see Chapter II.

3.1. Results

3.1.1. Switching and clustering in the ‘Category fluency test’

In this section, the two components that determine successful performance on a category fluency paradigm – clustering and switching (e.g., Troyer, 2000) – will be analysed for the CI subjects and their matched controls. The clustering measure reflects the ability to generate words within a subcategory, while the switching measure refers to the ability to shift to another subcategory when one cannot retrieve more words from the previous subcategory. Deficits in these two components may represent, in fact, semantic search problems or impaired access to semantic knowledge.

Table 1: Switching and clustering scores (means; standard deviations in parentheses) per category on the ‘Category fluency test’, by the CI subjects and the mNC subjects.

<table>
<thead>
<tr>
<th></th>
<th>CI subjects (n=65)</th>
<th>mNC subjects (n=65)</th>
<th>paired t test:</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of correct animals</td>
<td>12.80 (3.78)**</td>
<td>15.06 (4.21)**</td>
<td>t=3.320, p=.001</td>
</tr>
<tr>
<td>number of different subcategories of animals ('switching')</td>
<td>4.17 (1.33)</td>
<td>4.57 (1.31)</td>
<td>t=1.718, p=.091</td>
</tr>
<tr>
<td>number of correct animals per subcategory ('clustering')</td>
<td>3.24 (1.08)</td>
<td>3.52 (1.54)</td>
<td>t=1.063, p=.292</td>
</tr>
<tr>
<td>number of correct occupations</td>
<td>9.11 (2.76)**</td>
<td>11.17 (3.97)**</td>
<td>t=3.627, p=.001</td>
</tr>
<tr>
<td>number of different subcategories of occupations ('switching')</td>
<td>4.94 (1.47)</td>
<td>5.20 (1.34)</td>
<td>t=1.136, p=.260</td>
</tr>
<tr>
<td>number of correct occupations per subcategory ('clustering')</td>
<td>1.92 (.57)*</td>
<td>2.16 (.65)*</td>
<td>t=2.532, p=.014</td>
</tr>
</tbody>
</table>

Paired t test analyses showed that the mNC subjects generated significantly more correct responses in both the animal category and the occupation category. However, no significant differences were found between the mNC and the CI subjects in the number of different subcategories produced (i.e., no differences in switching ability) – either for the animals, though the difference was rather close to significance (p=.091), or for the occupations, as Figure 1 illustrates.
**Figure 1:** Absolute number of different subcategories produced (‘switching’ measure) in the ‘Category fluency test’, by each matched group separately.

Paired $t$ test analysis showed that the CI subjects and the mNC subject did not differ in the average number of correct animals named per subcategory (i.e., no difference in clustering ability in the animal category). However, the mNC subjects named significantly more correct occupations per subcategory than the CI subjects did (i.e., mNC subjects had a better clustering ability in the occupation category). Figure 2 illustrates these differences in clustering for both groups and both categories.

**Figure 2:** Proportion of correct exemplars per subcategory (‘clustering’ measure) in the ‘Category fluency test’, by each matched group separately.

In sum, the CI subjects generated fewer correct responses in both categories than their matched controls. In the animal category, the difference in production may be due to a slightly impaired switching ability in the CI subjects, as opposed to the mNC subjects, though the difference was not significant ($p=.091$). In the occupation category, the CI subjects showed significantly worse clustering than their controls, which may explain the impaired generation level in this category.
3.1.2. Frequency distribution of responses generated in the ‘Category fluency test’

In this section, the degree of integrity of semantic store will be investigated by examining the frequency distribution of responses generated by the subjects. It will be analysed whether the subjects generate relatively few uncommon (middle- and low-frequency) exemplars of a category, compared with common (high-frequency) exemplars. This allows for an investigation of the hierarchical organisation of semantic knowledge of the elderly subjects, which might indicate a bottom-up breakdown, as is frequently found in AD.

Table 2: Absolute and relative number of correct responses generated per frequency type (means; standard deviations in parentheses) per category on the ‘Category fluency test’, by the CI subjects and the mNC subjects.

<table>
<thead>
<tr>
<th>Category</th>
<th>CI subjects (n=65)</th>
<th>mNC subjects (n=65)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>number of HF</td>
<td>% to sum:</td>
</tr>
<tr>
<td>number of HF</td>
<td>(2.56)</td>
<td>=56.02%</td>
</tr>
<tr>
<td>animals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of MF</td>
<td>(2.73)**</td>
<td>=33.35%</td>
</tr>
<tr>
<td>occupations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of LF</td>
<td>(1.70)</td>
<td>=10.56%</td>
</tr>
<tr>
<td>animals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of LF</td>
<td>(1.60)*</td>
<td>=50.42%</td>
</tr>
<tr>
<td>occupations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of LF</td>
<td>(1.56)**</td>
<td>=28.30%</td>
</tr>
<tr>
<td>occupations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>number of LF</td>
<td>(1.59)</td>
<td>=21.17%</td>
</tr>
<tr>
<td>occupations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p<.01; * p<.05

Figures 3a and 3b: Distribution of absolute numbers of HF, MF and LF responses in the ‘Category fluency test’, by each matched group separately.
An ANOVA performed over the absolute number of responses generated in the animal category showed a significant group by frequency interaction effect ($F=4.703; df=2; p=.010$), as a result of the CI subjects naming significantly fewer MF responses than the mNC subjects did (paired $t$ test: $t=3.748; df=64; p=.000$; see also Figure 3a)$^{49}$. In addition, a slightly significant group by frequency interaction effect was found for the absolute number of responses generated in the occupation category ($F=3.286; df=2; p=.038$), due to the CI subjects naming significantly fewer HF ($t=-2.056; df=64; p=.044$) and MF responses ($t=3.533; df=64; p=.001$) than the mNC subjects did (see also Figure 3b)$^{50}$.

### 3.2. Discussion

The CI subjects showed, in comparison with the mNC subjects, less switching between subcategories in the animal category, though the difference was not significant ($p=.091$). Analysis of the frequency distribution of responses in the animal category showed that the lower number of correct responses of the CI subjects was primarily due to naming fewer MF exemplars than the mNC subjects did. The response profile of the CI subjects consisted of relatively many HF responses and relatively few MF responses, in comparison with the mNC subjects.

A slightly different pattern was found for the occupation category. The CI subjects demonstrated significantly less clustering in the occupation category. Furthermore, the lower number of correct responses of the CI subjects resulted from, particularly, naming fewer MF exemplars and also fewer HF responses than the mNC subjects did.

It may be concluded that the worse generation level in the animal category by the CI subjects results from a slightly less varied response profile (i.e., fewer different subcategories were generated – a worse switching ability), since the CI subjects produced a similar number of items per subcategory as the mNC subjects did (a similar clustering ability). In contrast to the pattern observed in the animal category, the lower number of occupations results from the production of fewer items per subcategory (worse clustering ability), since the CI subjects generated a similar number of different subcategories as the mNC subjects. An additional factor that may have caused the generally worse performance of the CI subjects may be the worse ability to name uncommon (MF) exemplars of the two categories presented. In addition, LF exemplars have added little to the total score, particularly in the animal category. However, this low number of LF exemplars may be expected in these subjects since they are able to name rather few exemplars in general.

In sum, it may be argued that the CI subjects have more difficulty retrieving MF exemplars, which may, in part, have led to their reduced generation level. In addition, worse switching and, particularly, worse clustering ability should be taken into account, despite the difference in pattern over the two categories. However, these conclusions must be regarded cautiously because of the experimental nature of the specific scores and the heterogeneous character of the CI group.

$^{49}$ Correspondingly, a Pearson Chi-Square Test demonstrated a significant effect of group by frequency ($\chi^2=8.725, df=2, p=.013$).

$^{50}$ Correspondingly, a Pearson Chi-Square Test demonstrated a slightly significant effect of group by frequency ($\chi^2=6.459, df=2, p=.040$).
CHAPTER V

4. Results and discussion of the second administration of the ‘Category fluency test’ – the development of scores over time

In this section, the various scores derived from the ‘Category fluency test’ will be examined for both administration periods (T₁ vs. T₂) and for various clinical subgroups, defined at T₂. In Chapter IV (section 4), four clinical subgroups were identified: 56 nondemented subjects that were classified at T₁ to the Normal Control group (the NonD-NC group), 28 nondemented subjects that were classified at T₁ to the Cognitively Impaired group (the NonD-CI group), 6 minimal dementia (Min.D) subjects, and 7 mild or moderate dementia (Mild/mod.D) subjects. It will be investigated whether the various scores show decline or a stable pattern of performance once subjects are diagnosed as demented at T₂ (i.e., mild or moderate dementia, consistent with the DSM-IV criteria of dementia), relative to the nondemented and the minimally demented subjects. In section 4.1.1, the development of switching and clustering scores over time will be examined for the different clinical subgroups. In addition, in section 4.1.2, the frequency distributions of responses over time will be investigated for these subgroups.

4.1. Results

4.1.1. Switching and clustering over time

In this section, the development of the two components that determine successful performance on a category fluency paradigm – clustering and switching (e.g., Troyer, 2000) – will be analysed for the different clinical subgroups. GLM Repeated Measures analyses were performed over the sum of correct responses produced for each category, the switching score and the clustering score, with within-subjects factor ‘administration period’ and between-subjects factor ‘clinical subgroup’. The clustering measure reflects the ability to generate words within a subcategory, while the switching measure refers to the ability to shift to another subcategory when one cannot retrieve more words from the previous subcategory. Deficits in these two components may represent, in fact, semantic search problems or impaired access to semantic knowledge.

Table 3: Switching and clustering scores (means; standard deviations) per category, per administration period (T₁ and T₂) on the ‘Category fluency test’, for each clinical subgroup separately.

<table>
<thead>
<tr>
<th>variable:</th>
<th>NonD-NC (n=56)</th>
<th>NonD-CI (n=28)</th>
<th>Min.D (n=6)</th>
<th>Mild/mod.D (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T₁:</td>
<td>T₂:</td>
<td>T₁:</td>
<td>T₂:</td>
</tr>
<tr>
<td>correct animals</td>
<td>15.66</td>
<td>15.55</td>
<td>14.29</td>
<td>13.75</td>
</tr>
<tr>
<td></td>
<td>(3.59)</td>
<td>(4.31)</td>
<td>(3.11)</td>
<td>(3.35)</td>
</tr>
<tr>
<td>’switching’</td>
<td>4.77</td>
<td>4.59</td>
<td>4.50</td>
<td>4.46</td>
</tr>
<tr>
<td>animals</td>
<td>(1.25)</td>
<td>(1.02)</td>
<td>(1.17)</td>
<td>(0.96)</td>
</tr>
<tr>
<td>’clustering’</td>
<td>3.54</td>
<td>3.46</td>
<td>3.26</td>
<td>3.15</td>
</tr>
<tr>
<td>animals</td>
<td>(1.60)</td>
<td>(0.93)</td>
<td>(0.66)</td>
<td>(0.71)</td>
</tr>
<tr>
<td>correct</td>
<td>11.61</td>
<td>12.20</td>
<td>10.36</td>
<td>10.50</td>
</tr>
<tr>
<td>occupations</td>
<td>(3.99)</td>
<td>(4.32)</td>
<td>(2.57)</td>
<td>(2.49)</td>
</tr>
<tr>
<td>’switching’</td>
<td>5.21</td>
<td>5.64</td>
<td>5.18</td>
<td>5.00</td>
</tr>
<tr>
<td>occupations</td>
<td>(1.17)</td>
<td>(1.59)</td>
<td>(1.39)</td>
<td>(1.25)</td>
</tr>
<tr>
<td>’clustering’</td>
<td>2.23</td>
<td>2.23</td>
<td>2.10</td>
<td>2.19</td>
</tr>
<tr>
<td>occupations</td>
<td>(.65)</td>
<td>(.70)</td>
<td>(.59)</td>
<td>(.61)</td>
</tr>
</tbody>
</table>

169
The number of correct responses produced – per category, administration period and clinical subgroup – is displayed in Figures 4a and 4b. In general, a stable pattern from T1 to T2 may be noted (no significant effect for ‘administration period’ was found: animals: $F=2.033$, $df=1, p=.157$; occupations: $F=2.445$, $df=1, p=.121$). Most clinical subgroups performed on a similar level at T2 as at T1, although, particularly, the Mild/mod.D subjects showed a decline of performance. Close to significant interaction effects were found between ‘administration period’ and ‘clinical subgroup’ (animals: $F=2.562$, $df=3, p=.060$; occupations: $F=2.295$, $df=3, p=.083$).

Both in the animal and in the occupation category, a significant effect for ‘clinical subgroup’ was found ($F=12.281$, $df=3, p=.000$; $F=8.443$, $df=3, p=.000$). In the animal category, Bonferroni tests showed that the NonD-NC group performed significantly better than the Min.D and the Mild/mod.D group ($p=.012$ and .000), while the NonD-CI performed better than the Mild/mod.D group ($p=.001$). In the occupation category, the NonD-NC group as well as the NonD-CI group produced significantly more correct responses than the Mild/mod.D group ($p=.000$ and .004). See also Figures 4a and 4b for an illustration of the differences in performance between the clinical subgroups.

**Figures 4a and 4b:** Number of correct responses produced in the ‘Category fluency test’ by each clinical subgroup.

The switching scores – per category, administration period and clinical subgroup – are displayed in Figures 5a and 5b. Again, a generally stable pattern from T1 to T2 may be noted, except for the Mild/mod.D subjects who showed declining switching scores in both categories, and the Min.D subjects who showed a modest decline in the occupation category (significant effect of ‘administration period’: animals: $F=5.485$, $df=1, p=.021$; occupations: $F=6.447$, $df=1, p=.013$; significant ‘administration period’

In addition, in both categories, a significant effect was found for ‘clinical subgroup’ ($F=8.370$, $df=3$, $p=.000$; $F=5.071$, $df=3$, $p=.003$). In the animal category, Bonferroni tests showed that the NonD-NC group performed significantly better than the Min.D group ($p=.024$), and the Mild/mod.D group ($p=.000$), while the NonD-CI group performed better than the Mild/mod.D group ($p=.003$). In the occupation category, the Mild/mod.D group performed significantly worse than the NonD-NC group ($p=.001$) and the NonD-CI group ($p=.027$). See also Figures 5a and 5b for an illustration of the differences in performance between the clinical subgroups.

**Figures 5a and 5b:** ‘Switching’ scores of the ‘Category fluency test’ (number of different subcategories) by each clinical subgroup.

GLM Repeated Measures analyses, performed over the *clustering scores*, only found a significant effect of ‘clinical subgroup’ in the occupation category ($F=4.067$, $df=3$, $p=.009$). This effect was caused by the Mild/mod.D group performing worse than the NonD-NC group ($p=.007$) and the NonD-CI group ($p=.036$). All the other effects were clearly non-significant: e.g., no evident change from $T_1$ to $T_2$ in general and no clear patterns for specific clinical subgroups, except for a slight improvement by the Min.D subjects in the animal category and for the Mild/mod.D subjects in the occupation category (see Figures 6a and 6b for an illustration).
Figures 6a and 6b: 'Clustering' scores of the 'Category fluency test' (number of correct responses per subcategory) by each clinical subgroup.

In sum, both nondemented groups performed rather stable from T1 to T2 on each score. In addition, the Min.D subjects slightly improved their production rate in the animal category, which may be explained by a larger cluster size (i.e., the switching score remained stable from T1 to T2). However, in the occupation category, the Min.D subjects showed a slightly reduced production rate, which may be explained by an impaired switching ability (i.e., the clustering score remained stable). Furthermore, the Mild/mod.D subjects generated fewer exemplars from T1 to T2, particularly in the animal category, which may be explained by a decreased switching ability. Their clustering score remained stable in the animal category, while it even improved a little in the occupation category.

4.1.2. Frequency distributions of responses over time
In this section, developments from T1 to T2 in the degree of integrity of semantic store will be investigated by examining the frequency distribution of responses generated by the subjects from the various clinical subgroups. It will be analysed whether the subjects generate relatively few uncommon (middle- and low-frequency) exemplars of a category, compared with common (high-frequency) exemplars. This allows for an investigation of the hierarchical organisation of semantic knowledge of the elderly subjects, which might indicate a bottom-up breakdown, as is frequently found in AD.
Table 4: Absolute and relative number of correct responses generated per frequency type (means; standard deviations) per category (animals and occupations), per administration period (T₁ and T₂) and per clinical subgroup on the ‘Category fluency test’.

<table>
<thead>
<tr>
<th>variable:</th>
<th>NonD-NC (n=56)</th>
<th>NonD-CI (n=28)</th>
<th>Min.D (n=6)</th>
<th>Mild/mod.D (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T₁:</td>
<td>T₂:</td>
<td>T₁:</td>
<td>T₂:</td>
</tr>
<tr>
<td>HF animals:</td>
<td>7.73 (2.28)</td>
<td>7.46 (2.38)</td>
<td>6.89 (2.23)</td>
<td>6.73 (2.64)</td>
</tr>
<tr>
<td>% to sum correct:</td>
<td>50.50% 49.41%</td>
<td>51.01% 51.58%</td>
<td>55.92% 64.99%</td>
<td>61.84% 81.77%</td>
</tr>
<tr>
<td>MF animals:</td>
<td>6.43 (2.40)</td>
<td>5.93 (2.61)</td>
<td>4.96 (2.35)</td>
<td>3.17 (1.83)</td>
</tr>
<tr>
<td>% to sum correct:</td>
<td>40.42% 36.94%</td>
<td>36.35% 35.39%</td>
<td>30.58% 25.68%</td>
<td>30.92% 17.13%</td>
</tr>
<tr>
<td>LF animals:</td>
<td>1.46 (1.44)</td>
<td>2.16 (1.78)</td>
<td>1.89 (1.79)</td>
<td>0.67 (.82)</td>
</tr>
<tr>
<td>% to sum correct:</td>
<td>8.90% 13.65%</td>
<td>12.64% 13.04%</td>
<td>13.50% 9.32%</td>
<td>7.24% 1.10%</td>
</tr>
<tr>
<td>HF occupations:</td>
<td>5.20 (1.82)</td>
<td>5.16 (1.98)</td>
<td>4.50 (2.08)</td>
<td>5.17 (1.83)</td>
</tr>
<tr>
<td>% to sum correct:</td>
<td>47.11% 44.84%</td>
<td>46.84% 43.06%</td>
<td>53.01% 49.16%</td>
<td>58.53% 70.00%</td>
</tr>
<tr>
<td>MF occupations:</td>
<td>4.02 (2.48)</td>
<td>3.66 (2.01)</td>
<td>3.75 (1.62)</td>
<td>2.67 (1.82)</td>
</tr>
<tr>
<td>% to sum correct:</td>
<td>32.91% 28.93%</td>
<td>28.96% 36.50%</td>
<td>25.44% 31.19%</td>
<td>32.62% 17.14%</td>
</tr>
<tr>
<td>LF occupations:</td>
<td>2.39 (1.87)</td>
<td>3.38 (2.19)</td>
<td>2.25 (1.60)</td>
<td>2.33 (1.63)</td>
</tr>
<tr>
<td>% to sum correct:</td>
<td>19.98% 26.23%</td>
<td>23.95% 20.44%</td>
<td>21.55% 19.64%</td>
<td>8.85% 12.86%</td>
</tr>
</tbody>
</table>

GLM Repeated Measures analyses were performed over the absolute number of responses, first, for the animal category, and, secondly, for the occupation category, with within-subjects factors ‘administration period’ and ‘frequency’ and between-subjects factor ‘clinical subgroup’.

Figures 7a and 7b present the distribution of responses according to frequency type generated in the animal category. In general, the NonD-NC group performed significantly better than the Min.D and the Mild/mod.D group, while the NonD-CI group performed significantly better than the Mild/mod.D group (significant effect ‘clinical subgroup’: $F=12.307, df=3, p=.000$; Bonferroni post-hoc tests). In addition, it may be noted that both the Min.D and the Mild/mod.D group generated rather few MF exemplars, relative to the NonD-NC and the NonD-CI subjects (significant ‘frequency’ by ‘clinical subgroup’ interaction effect: $F=2.915, df=6, p=.010$; see also Figures 7a and 7b: the difference between the (minimally) demented subjects and the nondemented subjects is greatest in the MF condition). Focusing on the specific administration periods, it may be noted that at T₁, the distribution of the HF, MF and LF responses was identical for the Min.D and the Mild/mod.D subjects. At T₂, the Min.D subjects showed a similar performance level as at T₁, though they generated slightly more HF responses at T₂. However, the Mild/mod.D subjects showed decline of performance at T₂, characterised by fewer responses in each frequency category, which led to a decrease of the percentage of MF and the LF responses (i.e., no LF responses were generated) and an increase of the percentage of HF responses (see
Table 4; though a clearly lower absolute number of HF responses was generated). The almost significant ‘administration period’ by ‘clinical subgroup’ interaction effect ($p=.058$) supports these observations (see Figures 7a and 7b).

Figures 7a and 7b: Distribution of absolute numbers of HF, MF and LF animals in the ‘Category fluency test’, at T₁ and T₂, by each clinical subgroup.

Figures 8a and 8b present the distribution of responses according to frequency type generated in the occupation category. In general, both the NonD-NC and the NonD-CI group performed significantly better than the Mild/mod.D group (significant effect ‘clinical subgroup’: $F=8.464$, $df=3$, $p=.000$; Bonferroni post-hoc tests). In addition, from Figures 8a and 8b it may be noted that the relatively low number of MF exemplars observed in the Min.D and the Mild/mod.D subjects in the animal category is less obvious in the occupation category. Particularly at T₁, the distribution of responses does not differ greatly between the four clinical subgroups. The only exceptions may be the NonD-NC subjects who generated rather many MF responses, and the Mild/mod.D subjects who generated clearly fewer LF responses than the other clinical subgroups. At T₂, the NonD-NC and the NonD-CI group performed in a similar manner as at T₁. However, both the Min.D and the Mild/mod.D group showed decline of performance, particularly the Mild/mod.D subjects (see Table 4 and Figure 8b). The Min.D subjects generated fewer HF responses and slightly fewer LF responses. The Mild/mod.D subjects generated fewer HF and fewer MF responses; their LF generation level was at T₁ already close to 0. These observations were supported by a close to significant ‘administration period’ by ‘clinical subgroup’ interaction effect ($p=.080$).
Figures 8a and 8b: Distribution of absolute numbers of HF, MF and LF occupations in the ‘Category fluency test’, at T1 and T2, by each clinical subgroup.

4.2. Discussion

In sum, both the Min.D and the Mild/mod.D group have a more or less stable generation level from T1 to T2. However, in the occupation category, the NonD-NC subjects show a slightly improved generation rate, which may be explained by the increase of LF responses. Furthermore, the Min.D subjects show a slightly improved generation rate in the animal category, which may be explained by a larger cluster size, consisting of more HF responses, relative to their performance pattern at T1. However, in the occupation category, the Min.D subjects produce fewer responses at T2, which may be explained by their impaired switching ability and mainly results in fewer HF responses. In addition, the Mild/mod.D subjects show a clear decrease of generation level, particularly in the animal category, which may be caused by a reduced switching ability and leads to fewer HF, MF and LF responses. In addition, it may be noted that their response profile is mainly characterised by HF responses; they hardly produce LF responses and the number of MF responses has also dropped considerably at T2.

Combining the response characteristics of the Min.D and the Mild/mod.D group, the decreased switching ability in both groups may be noted, rather than a poor clustering ability, while officially demented subjects perform worse than minimally demented subjects. In addition, both the Min.D and the Mild/mod.D group are characterised by a relatively small number of MF responses, specifically in the animal category, while the pattern is again more pronounced in the Mild/mod.D subjects. The pattern of a worse ability to name less common (MF and LF) responses deteriorates at T2, but only in the Mild/mod.D subjects. Thus, only when the subjects develop into a more advanced stage of dementia, this pattern may be detected. In addition, the Mild/mod.D subjects generate fewer common (HF) exemplars at T2, which increases the performance difference with the Min.D subjects and, specifically, with the nondemented subjects.
Within the nondemented group, the NonD-CI subjects perform generally worse than the NonD-NC subjects. However, the pattern of the CI subjects' specifically naming fewer MF exemplars than their controls (see section 3 of this chapter) cannot be detected in the current section, where the T2 (minimally) demented subjects are excluded from the NonD-CI group. Thus, this specific pattern is restricted to subjects that are in a (pre)clinical stage of dementia. The implications of these findings for the theory concerning semantic search (or access) problems and the integrity of semantic memory store will be discussed in section 6.

5. Predictive value of the various scores of the ‘Category fluency test’ regarding dementia

As was described in the introduction, Weingartner, Kawas, Rawlings and Shapiro (1993) found that normal elderly controls generated more uncommon exemplars from semantic categories than AD patients did 2½ years prior to the presumed onset of AD. In this section, the T1 performance of the 10 subjects that were diagnosed demented at T2 (i.e., mild or moderate dementia according to CAMDEX) and 10 for age, education and sex matched NC subjects (see Chapter II for the matching criteria) will be compared by means of paired t test analyses. In addition, a discriminant analysis will determine which specific fluency scores are best able to distinguish the T2 demented subjects from the subjects that did not receive a dementia diagnosis at T2.

5.1. Results

Paired t test analyses were performed over the various (T1) scores that were derived from the ‘Category fluency test’ in this Chapter51. Differences between the 10 subjects that were diagnosed demented at T2 and 10 matched NC subjects were tested for significance. The variables that showed (close to) significant differences between the two groups are presented in Table 552.

Table 5: T1 scores (means; standard deviations) derived from the ‘Category fluency test’ that led to (close to) significant differences between the T2 demented subjects (n=10) and age-,education- and sex-matched NC subjects.

<table>
<thead>
<tr>
<th>‘Fluency’ variable:</th>
<th>T2 demented ss.</th>
<th>matched NC ss.</th>
<th>paired t test:</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct animals</td>
<td>11.70 (4.24)</td>
<td>15.80 (3.39)</td>
<td>t=3.188, p=.011*</td>
</tr>
<tr>
<td>correct occupations</td>
<td>7.80 (3.39)</td>
<td>11.80 (3.39)</td>
<td>t=3.833, p=.004**</td>
</tr>
<tr>
<td>‘switching’ occupations</td>
<td>4.70 (1.34)</td>
<td>5.80 (1.40)</td>
<td>t=2.181, p=.057</td>
</tr>
<tr>
<td>‘clustering’ occupations</td>
<td>1.62 (.51)</td>
<td>2.10 (.63)</td>
<td>t=2.319, p=.046*</td>
</tr>
<tr>
<td>MF animals</td>
<td>4.10 (2.51)</td>
<td>6.90 (2.73)</td>
<td>t=2.662, p=.026*</td>
</tr>
<tr>
<td>MF occupations</td>
<td>2.40 (1.43)</td>
<td>4.80 (2.44)</td>
<td>t=2.676, p=.025*</td>
</tr>
</tbody>
</table>

51 For both the animal and the occupation category: number of correct responses, number of intrusions, number of repetition-type perseverations, number of same exemplar-type perseverations, ‘switching’ score, ‘clustering’ score, number of HF responses, number of MF responses, and the number of LF responses.

52 Comparing the subjects that received a minimal, mild or moderate dementia diagnosis at T2 (n=16) and their matched controls — i.e., classifying the T2 minimal dementia subjects to the ‘demented’ group, as was also done in several analyses in Chapter IV — did not show different results, except for the non-significance of the ‘switching’ and ‘clustering’ scores in the occupation category (p=.155 and .053).
In addition, a ‘stepwise’ discriminant analysis was performed over the various $T_1$ scores. Only one score came forward from the analysis as a discriminating variable between the 10 $T_2$ demented and 125 nondemented subjects$^{53}$: the number of correct occupations (Residual variance = .857; Exact $F = 6.173; df=133; p = .014$). The second variable (the number of repetition-type perseverations of the occupation category) did not significantly minimise the sum of the remaining variance; significance of $F$ to enter was .082. The classification of the subjects to the demented or the nondemented group, according to various fluency variables, is displayed in Table 6. Classification has been performed with an equal probability (50%) for subjects being classified to the demented or the nondemented group, as well as with the prior probability being computed from the respective group sizes.

Table 6: Classification of the subjects to the demented (mild or moderate dementia at $T_2$) or the nondemented group (no or minimal dementia at $T_2$)$^{54}$ and accuracy of prediction, resulting from discriminant analyses performed over various variables derived from the ‘Category fluency test’. Discrimination measure $d'$ is also calculated.

<table>
<thead>
<tr>
<th>variables</th>
<th>classification method*</th>
<th>accuracy of classification</th>
<th>proportion of ‘true positives’</th>
<th>proportion of ‘true negatives’</th>
<th>$d'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>none (baseline)</td>
<td>A</td>
<td>50%</td>
<td>5/10</td>
<td>62.5/125</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>86.7%</td>
<td>1/10</td>
<td>116/125</td>
<td>0.19</td>
</tr>
<tr>
<td>‘occupations correct’</td>
<td>A</td>
<td>63.0%</td>
<td>6/10</td>
<td>79/125</td>
<td>0.58</td>
</tr>
<tr>
<td>(‘stepwise’ analysis)</td>
<td>B</td>
<td>92.6%</td>
<td>0/10</td>
<td>125/125</td>
<td>0.00</td>
</tr>
<tr>
<td>‘clustering’ occupations</td>
<td>A</td>
<td>62.2%</td>
<td>7/10</td>
<td>77/125</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>92.6%</td>
<td>0/10</td>
<td>125/125</td>
<td>0.00</td>
</tr>
<tr>
<td>MF animals</td>
<td>A</td>
<td>63.7%</td>
<td>7/10</td>
<td>79/125</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>92.6%</td>
<td>0/10</td>
<td>125/125</td>
<td>0.00</td>
</tr>
<tr>
<td>MF occupations</td>
<td>A</td>
<td>67.4%</td>
<td>7/10</td>
<td>84/125</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>92.6%</td>
<td>0/10</td>
<td>125/125</td>
<td>0.00</td>
</tr>
<tr>
<td>all fluency scores together</td>
<td>A</td>
<td>84.4%</td>
<td>9/10</td>
<td>105/125</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>93.3%</td>
<td>3/10</td>
<td>123/125</td>
<td>1.52</td>
</tr>
</tbody>
</table>

*classification method:
A=all groups equal probability (50%)
B=prior probability computed from group size (nondemented group: 125/135→92.6%;
demented group: 10/135→7.4%)

$^{53}$ 12 subjects were excluded in these analyses: the 11 subjects that were classified as minimally demented according to CAMDEX at $T_1$, while they were classified as nondemented at $T_2$, as well as the subject that was classified as minimally demented at $T_1$ and of whom no information was available at $T_2$ (see also Chapter IV and Appendix E in Chapter II). These subjects were also excluded in the previous analyses in section 4 and 5 of the present chapter.

$^{54}$ Classifying the minimally demented subjects to the ‘demented’ group or excluding them (as was also done in Chapter IV) did not show different results, except for the significance of the ‘animals correct’ score in the ‘stepwise’ analysis, instead of the ‘occupations correct’ score, when classifying the minimally demented subjects to the ‘demented’ group (Residual Variance=.852, $F=9.823, df=133$, $p=.002$; 65.9% accuracy of prediction, $d'=0.47$ (method A); 88.9%, $d'=0.77$ (method B)).
**Figure 9a:** The accuracy of prediction regarding (mild or moderate) dementia of the 'occupations correct' score, the 'MF occupations' score and all fluency scores together at T1, relative to the other variables derived from the memory test battery (see Chapter IV, section 5).

![Accuracy of prediction (%)](image)

- ■ both groups equal prob.
- □ prior prob. from group size

**Figure 9b:** Discrimination measure $d'$, regarding the differentiation between the demented and the nondemented group.

![Discrimination $d'$](image)

- ■ $d'$ (both groups equal prob.)
- □ $d'$ (prior prob. from group size)
5.2. Discussion
The demented subjects generated significantly fewer MF exemplars in both the animal and the occupation category than matched controls, two years before the diagnosis of dementia according to DSM-IV was made. This finding corresponds with the results of Weingartner et al. (1993), although no significant difference in the number of LF exemplars was found in the current study (i.e., LF responses were hardly ever generated). However, the number of MF exemplars was not selected by the ‘stepwise’ discriminant analysis as the best predictor of dementia. Only the ‘occupations correct’ score came forward from the analysis – the addition of other fluency variables did not significantly improve the prediction. Nevertheless, the predictive value (% of correct classifications) and the discriminative value (d') of MF exemplars (whether in the animal or the occupation category) were slightly better than the respective values of the ‘occupations correct’ score (see the grey cells in Table 6). It may be noted that both variables were highly correlated: the sum of correct exemplars in both the animal and the occupation category was most highly correlated with the respective number of MF exemplars generated, relative to the number of HF and LF exemplars: Pearson coefficients .763 and .688). This explains why the addition of these variables did not significantly improve the prediction of dementia.

Nonetheless, it may be noted from the prediction according to the ‘occupations correct’ score (or the number of MF exemplars; classification method A) that many subjects were falsely classified to the demented group (i.e., a low specificity of 63.2 to 67.2%) and few demented subjects were correctly identified (i.e., a low sensitivity of 60 to 70%). Though all fluency scores together lead to a more accurate prediction than the ‘occupations correct’ score alone (i.e., evidently better sensitivity and specificity), the predictive variables described in Chapter IV (section 5) still lead to the best prediction of dementia (see also Figures 9a and 9b). When the group size of the demented subjects is considered in the classification process (method B), hardly any difference in accuracy of classification with the variables described in Chapter IV is found (see Figure 9a). However, the d' values of the ‘occupations correct’ score and the ‘MF occupations’ score are equal to 0, which is caused by classifying all subjects to the ‘nondemented’ group (a sensitivity level of 0; see Table 6). In addition, the d' value of all fluency scores together is lower than in most other variables of the memory test battery (see Figure 9b), which is caused by a sensitivity level of only 30%, just above the baseline level.

Thus, it may be argued that the differences in performance characteristics between the nondemented and the demented subjects (see section 4) may only be large enough at T2, rather than at T1, the preclinical stage. Therefore, it may be concluded that these scores do not improve the prediction of dementia, two years before the diagnosis was made. Only when the disease is in a clinical stage, the differences in characteristics of fluency performance between demented and nondemented subjects are more pronounced.

6. Conclusion
In sum, the development from T1 to T2 of the various scores derived from the ‘Category fluency test’ is characterised by a rather stable performance in the NonD-NC and the NonD-Cl subjects. The subjects that originated from the NC group perform better than the subjects that were classified to the CI group, but except for
this quantitative difference, the performance patterns are qualitatively similar. The Min.D and, particularly, the Mild/mod.D subjects show a decline of performance. This decline leads to a lower generation level, which is mainly explained by a decreased ability to switch from one subcategory to another. In addition, the response profile of the Min.D and, particularly, the Mild/mod.D subjects is characterised by a relatively small number of MF responses. This pattern intensifies at T2, but only in the Mild/mod.D subjects. Thus, when these subjects are in a clinical stage of dementia, their response profile predominantly consists of highly frequent exemplars; they rarely generate less common exemplars. This finding refers to the bottom-up breakdown of semantic knowledge, described in the introduction.

The finding of a worse ability to name less common exemplars, as well as the reduced switching ability of the demented subjects, corresponds with results found in studies testing AD patients with the same or other semantic categories (e.g., Tröster, Salmon, McCullough & Butters, 1989; Chertkow & Bub, 1990; Weingartner, Kawas, Rawlings & Shapiro, 1993; Beatty, Testa, English & Winn, 1997). However, the 'clustering' measure (e.g., Troyer, 2000) did not clearly differentiate between the demented subjects (in their clinical stage) and the nondemented subjects. This finding corresponds with a study by Testa et al. (1998), in which supermarket items were used. Testa et al. reported that their AD patients showed reduced cluster sizes when they were moderately or severely demented; cluster size for mild AD patients was normal.

Nonetheless, it may be concluded that the subjects in the current study that are clinically demented suffer from semantic search problems and a deteriorated semantic memory store (e.g., Chertkow & Bub). Both aspects of fluency performance lead to an evident difference with the at T2 nondemented subjects and show a declined pattern of performance from T1 to T2. The semantic search problems are characterised by switching difficulties, rather than by a worse clustering ability. The deteriorated semantic memory store is characterised by the impaired ability to name less common exemplars of the category presented. Thus, the clinically demented subjects (and to a lesser degree the minimally demented subjects) produce only a few subcategories, mainly consisting of the most common exemplars, while the nondemented subjects produce additional subcategories, which also include less common exemplars.

Concerning the predictive value of the various fluency scores, it may be noted that a few T1 scores differentiated significantly between the at T2 demented subjects and matched controls (i.e., the sum of correct responses and the number of MF exemplars in both the animal and the occupation category, and the 'cluster size' in the occupation category). Thus, the findings by Weingartner et al. (1993) were replicated: normal elderly controls generated more uncommon (MF) exemplars from semantic categories than demented subjects did two years prior to the presumed onset of the disease. However, the predictive value of this variable relative to other (memory) scores, was rather disappointing. In addition, the difference in predictive value between the number of MF exemplars and the sum of correct responses in general is only minor. Thus, it makes little difference whether to choose for the one or the other score (note that both scores are highly correlated). Nonetheless, the score indicating the number of MF responses provides us with more information concerning the nature of performance of the (pre)clinically demented subjects: it explains the deficits found on this test. These deficits are characterised by a reduced availability of less common exemplars of semantic networks and may be described by a bottom-up breakdown of semantic knowledge. However, from a practical point of view, it may be concluded that the extra scores derived from the 'Category fluency test', as opposed to, for
example, the ‘Paired-associate learning test’, are better classification aids when dementia has already been diagnosed than when it is still in a preclinical stage.
Appendix A: Hierarchical model of exemplars of animals named in the ‘Category fluency test’ (partially based on Stichmann, Krettschmar, Keijl & Nieuwenkamp (1999) and http://www.dierentuin.net/database.html); each exemplar notes the number of times it was named by all normal control subjects tested at T1 (n=82; mNC and nNC), excluding perseverative answers, overlapping ‘labels’ printed in italics:

![Diagram of animal categories

animals

- pets & farm animals
- wild animals
- predators
- reptiles & amphibians
- molluscs
- insects
- birds
- fish

HF
- cat (64)
- dog (64)
- cow (47)
- horse (42)
- mouse (34)
- pig (27)
- sheep (25)
- donkey (22)
- rabbit (21)
- chicken (20)
- goat (17)
- cock (7)
- cavia (7)
- hamster (5)
- marmot (4)
...and others...

HF
- elephant (34)
- monkey (34)
- rat (29)
- deer (16)
- zebra (11)
- giraffe (10)
- hare (9)
- elk (9)
- roe (9)
- fox (7)
- seal (7)
- squirrel (6)
- camel (6)
- hedgehog (5)
- whale (5)
...and others...

MF
- tiger (47)
- lion (46)
- bear (17)
- panther (13)
- wolf (11)
- hyena (7)
- otter (6)
- leopard (5)
- weasel (5)
- polecat (3)
- puma (3)
- badger (3)
- polar bear (2)
- panda (2)
- lynx (2)
...and others...

MF
- snake (19)
- crocodile (11)
- frog (11)
- iguana (2)
- toad (2)
- chameleon (1)

MF
- worm (8)
- louse (7)
- mussel (2)
- oyster (1)
- cuttlefish (1)

MF
- ant (14)
- parrot (16)
- mosquito (6)
- flea (6)
- fly (6)
- beetle (4)
- wasp (3)
- bee (3)
- butterfly (2)
- caterpillar (2)
- ladybird (2)
- scorpion (2)
- cockroach (2)
...and others...

MF
- bird (27)
- pike (6)
- duck (15)
- sparrow (14)
- canary (14)
- pigeon (11)
- heron (8)
- magpie (8)
- crow (7)
- starling (5)
- swallow (5)
- parakeet (5)
...and others...

LF
- fish (22)
- shark (4)
- plaice (3)
- codfish (3)
- perch (3)
- flounder (2)
- herring (2)
...and others...
Appendix B: Hierarchical model of exemplars of occupations named in the ‘Category fluency test’ (partially based on Centraal Bureau voor de Statistiek, 1996, 1999); each exemplar notes the number of times it was named by all normal control subjects tested at T1 (n=82; mNC and nNC), excluding perseverative answers, overlapping ‘labels’ printed in italics: