Intact Primary Memory in Mild to Moderate Alzheimer Disease: Indices from the California Verbal Learning Test*

Eileen Simon1, Larry Leach1, Gordon Winocur2, and Morris Moscovitch1,3

1Baycrest Centre for Geriatric Care, 2Rotman Research Institute, Baycrest Centre for Geriatric Care and Trent University, and 3Rotman Research Institute, Baycrest Centre for Geriatric Care and University of Toronto, Erindale College

ABSTRACT

The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) was administered to patients with mild to moderate Alzheimer disease (AD) (Group AD; n=13) and to a control group of normal older adults (Group NC; n=13) matched on age and education. Two measures were used to determine whether primary memory (PM) is impaired in early AD. One measure, considered a relatively “pure” measure of PM, is based on the procedure developed by Tulving and Colotla (1970) which considers an item to be recalled from PM if no more than six items intervene between its presentation and recall. The other measure is the more commonly used recall from recency. No significant difference between the AD and NC Groups was found, both on the Tulving and Colotla measure, as well as on the recall from recency measure of PM. A significant difference was obtained on two measures of secondary memory (SM), namely, Tulving and Colotla’s measure and recall from the primacy and middle regions of the list of words. In comparison to NC, the AD patients showed little evidence of learning over the five trials, and poor retention even over short delays. In addition, the patients with AD showed deficits in clustering words by taxonomic category at recall. We conclude that impairment in PM cannot be used as a diagnostic marker of AD in the early stages of the disease process.

The information-processing approach delineates two key components underlying memory functioning. Primary memory (PM), more commonly known as short-term memory, is immediate memory for information still in mind. Secondary memory (SM), or long-term memory, is delayed memory for newly learned information. The distinction between PM and SM (e.g., Waugh & Norman, 1965) has been useful in the study of normal age-related changes in memory, as well as in distinguishing among different memory disorders (for reviews see Craik & Jennings, 1992; Kaszniak, Poon, & Riege, 1986). For example, amnesic patients with focal lesions to the medial temporal lobe and related structures have impaired SM but normal PM, whereas some patients with conduction aphasia show the reverse pattern. Memory impairment is also usually the first, most noticeable symptom in patients with Alzheimer Disease (AD), its primary distinguishing feature being a difficulty in learning new information. The question of interest is whether the initial memory loss observed in AD is attributable only to a disorder of SM or whether PM is also affected. In AD, there is conclusive evidence from clinical observations and research that there is a deficit in SM (see summary of studies in Kaszniak et al., 1986 and a review by Morris & Baddeley, 1988). This deficit appears in the earliest stages

* This study was supported by funds from the Medical Research Council of Canada to G. Winocur and the Ontario Mental Health Foundation to M. Moscovitch. We gratefully acknowledge the assistance of Margaret Newson in the data collection and analysis, Edith Kaplan for her support and insight on the California Verbal Learning Test, and Morris Freedman for referring the patients. Address correspondence to: Eileen Simon, Department of Psychology, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, North York, Ontario, Canada, M6A 2E1.

Accepted for publication: August 10, 1993.
of the disease and therefore plays a key role as a behavioural indicator to support a clinical diagnosis. The evidence regarding the status of PM and its contribution to the observed deficiency in SM is much less conclusive. It is this issue that we address in our paper.

A variety of techniques have been used to assess PM in patients with AD. The most commonly used tests include forward digit span (Corkin, 1982; Storandt, Botwinick, Danziger, Berg, & Hughes, 1984), the Brown-Peterson distractor task (Corkin, 1982; Morris, 1986), and free recall from the recency part of a list of words (Miller 1971; Spinnler, Della Sala, Bandera, & Baddeley, 1988). Performance on these tests, however, often involves both PM and SM (e.g., Baddeley, 1976; Craik, 1977; Moscovitch, 1982a; Waugh & Norman, 1965), making it difficult to determine with certainty whether the deficit is associated with disorders in one or the other type of memory.

One test that attempts to take into account the multidimensional nature of memory, including the distinction between PM and SM, is the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987). By including several measures of the processes involved in memory, the CVLT has been used successfully to describe the memory deficits of patients suffering from a variety of disorders including, for example, head injury (Crosson, Novack, Trenever, & Craig, 1988), subcortical dementia (Massman, Delis, Butters, Levin, & Salmon, 1990), and AD (Kramer et al., 1988; Kramer, Levin, Brandt, & Delis, 1989).

The recommended procedure for assessing PM in the CVLT is the percent recency in free recall of a 16-item, categorized list of words. Percent recency is the proportion recalled from the last four items of the list, while recall from the balance of the list is a measure of SM. As noted earlier, however, recency may be contaminated by SM. Furthermore, both recency and primacy measures yielded modest to inadequate reliability coefficients in a recent study on the reliability of an alternate form of the CVLT (Delis et al., 1991). These are at least two reasons for considering alternative measures of PM.

A potentially better measure of PM, one that is considered to be a relatively "pure" measure (Moscovitch, 1982a; Watkins, 1975), was developed by Tulving and Colotla in 1970. According to Tulving and Colotla, recall is considered to be from PM if not more than six items occur between an item's presentation and its recall. The balance of recalled items is considered to come from SM.

Only a few studies involving patients with AD used the Tulving and Colotla measure of PM. Moscovitch (1982b) reports normal PM in patients in the early stages of AD. On the other hand, Wilson and his colleagues (Wilson, Bacon, Fox, & Kaszniak, 1983) found a deficit in PM. Their conclusion that PM is impaired in patients with AD, however, needs to be qualified for at least two reasons.

First and most importantly, the severity of the dementia in the AD group was not specified or likely controlled, and may have been the major factor contributing to the results. PM may have been deficient in the more advanced stages but still normal in patients with mild dementia who were in the early stages of the disease. We have no way of distinguishing the degree of severity from their data. Second, the presentation procedure may have penalized the AD patients. The task involved immediate free recall of four separate 12-word lists. Whereas PM improved in the normal controls, the patients' performance did not improve over trials. No analysis was presented comparing the performance of the patients and the normal controls on the first list where trial effects were absent.

A third study by Martin and his colleagues (Martin, Brouwers, Cox, & Fedio, 1985) examined PM and SM in patients with AD on the first recall trial of an eight-word list. The only significant effect was fewer total words recalled by patients with AD relative to normal controls. It is not clear from the analyses presented whether PM was impaired. Furthermore, the list used in this study was relatively short, and the level of severity in the AD group was not controlled in the design of the experiment. In light of the empirical limitations, as well as the po-
that PM is impaired in early AD.

The main objective of the present study was to determine whether there is indeed a PM deficit in mild to moderately impaired patients with AD. Two measures were used for determining PM: Tulving and Colotla’s (1970) procedure and the recency part of serial position in free recall (i.e., the last four items). By using both measures we also were able to compare the sensitivity of a commonly used clinical measure of PM with one that has been shown to be the most stable and reliable in an experimental or laboratory setting. Percent recency, which corrects for total words recalled, is the measure recommended in the instructions to the CVLT (Delis et al., 1987). In the present study, the mean number was used instead of percent because we were interested in the absolute level of recall from recency. In addition, percent recency is not appropriate for AD patients, who often recall only recent items, which would result in recall close to 100%. Other measures associated with recall in SM, such as semantic clustering and retention, were also examined to provide more information about the nature of the SM deficit in AD patients. The clustering measure provided insight into whether semantic organizational factors contribute to SM memory in AD. By measuring the effects of delay we would know whether information in SM was retained as well by AD as by normal people.

METHOD

Subjects

Twenty-six subjects were included in this study: 13 patients with probable AD (Group AD; 6 males, 7 females) and 13 normal controls (Group NC; 6 males, 7 females).

AD patients were diagnosed as having probable Alzheimer Disease based on neurological and neuropsychological evaluations. Only patients whose Global Deterioration Scale (Reisberg, Ferris, de Leon, & Crook, 1982, 1988) rating was mild (3) to moderate (4) were included.

Group NC consisted of community-dwelling volunteers recruited from the pool of subjects registered at the Memory Laboratory of Erindale College, University of Toronto. They were matched to the AD patients for age, education, and sex, and were paid a nominal sum for their participation. All control subjects signed a consent form before testing.

The mean ages (and standard deviations) of the NC and AD groups of subjects were 71.8 (6.9) and 72.5 (7.3) years, respectively. The mean years of education (and standard deviations) were, in order, 13.5 (3.6) and 11.2 (3.6). There was no statistically significant difference between the groups’ years of education, t(24) = 1.67, p = .108.

Measures

The data were obtained from each of the five immediate learning trials (Monday list) of the CVLT, as well as the two recall trials following a short (5 min) and long (20 min) filled delay. The learning trials comprise a 16-item shopping list from four different categories. A word never follows another from the same category. Five learning trials are presented orally in the same order, each followed by a free recall test. The measures include: mean recall on immediate learning trials, mean recall from recency (last four words), mean recall from middle (middle eight words), mean recall from primacy (first four words), a semantic cluster score, short delay recall versus Trial 5, and long delay recall versus Trial 5.

Semantic clustering is a measure of consecutive recall of words from the same taxonomic category, reflecting the organization that is imposed on the list of items. One point is given for each correct word that follows another one from the same category, providing an observed correct semantic clustering score. The ratio of observed and expected scores is then calculated to take into account the clustering that might occur by chance given the total number of words and the number of different categories recalled on a particular trial. In addition, we calculated the number of items recalled from PM and SM for each of the five trials, according to Tulving and Colotla (1970) in which an item is from PM if not more than six items intervene between its presentation and recall.

Procedure

The CVLT was administered routinely to patients as part of a neuropsychological assessment. The protocols for Group AD were retrieved from the files of the Department of Psychology, Baycrest Centre for Geriatric Care covering the period from January 1986 to December 1991. For Group NC the CVLT was administered to each subject individually according to the procedure recommended in the CVLT manual.
Fig. 1. Mean number of words recalled on immediate learning trials in NC and AD. (a) Mean total recalled. (b) Mean recall in primacy (first four words), middle (middle eight words), and recency (last four words). (c) Mean recall in PM and SM using Tulving and Colotla's (1970) measure.
RESULTS

The data were analyzed using a mixed factorial analysis of variance (ANOVA).

Immediate Recall
Mean recall of words for each group as a function of the five immediate learning trials is shown in Figure 1a. The AD patients obtained lower scores than NC beginning with the first trial, and showed minimal learning over trials. The performance of the NC subjects improved as learning progressed from the first to the fifth trial. These observations were confirmed by the analysis of variance; significant results were obtained for the main effects of groups, $F(1, 24) = 41.10, p \leq .001$, trials, $F(4, 96) = 40.30, p \leq .001$, and the interaction of groups by trials, $F(4, 96) = 3.14, p \leq .01$, reflecting the increasingly larger difference in recall between the two groups as learning increased from the first to the later trials.

The mean number of words recalled from primacy (the first four words), middle (middle eight words), and recency (the last four words) on each of the five immediate learning trials is shown in Figure 1b. As can be seen, performance in primacy and middle resulted in large differences between groups which increased on later trials. Statistical significance was obtained for the main effects of groups, $F(1, 24) = 41.10, p \leq .001$, and trials, $F(4, 96) = 40.30, p \leq .001$, and the interactions between groups and trials, $F(4, 96) = 3.14, p \leq .05$, and of memory and trials, $F(8, 192) = 2.69, p \leq .01$. On the other hand, the groups showed comparable performance in recall from recency. This was confirmed by the significant interaction between groups and type of memory, $F(2, 48) = 14.47, p \leq .001$.

Mean recall in PM and SM on Tulving and Colotla’s (1970) measure, is presented in Figure 1c. Observation of the data reveals a pattern of results similar to that obtained for recall from recency on the one hand, and primacy and middle on the other. There was no difference in recall performance between AD and NC in PM, whereas there was a significant difference in SM. This was confirmed by the analysis of variance; in addition to main effects and two-way interactions, a significant effect was obtained on the interaction between groups, type of memory, and trials, $F(4, 96) = 4.34, p \leq .01$.

As noted in the CVLT manual, the results of Trial 1 provide a good measure of recency because they are not contaminated by any trial effects. Mean recall from recency on Trial 1 was identical for the AD and NC groups, namely, 1.31. Using Tulving and Colotla’s measure of PM, the means for AD and NC on Trial 1 were 1.54 and 1.23, respectively. Post-hoc analyses (Newman-Keuls) were performed on these data and none was significant. Mean recall on Trial 1 from primacy and middle was .54 and 1.15 for AD, and 1.92 and 3.54 for NC. Using Tulving and Colotla’s measure of SM, the means for AD and NC were 1.46 and 5.64. Post-hoc analyses comparing these means were all significant ($p \leq .01$).

Thus, the AD patients showed no evidence of impairment in PM on two different measures, namely, mean recall in recency and Tulving and Colotla’s (1970) measure of PM. On the other hand, AD patients were impaired on both measures of SM. This differential pattern of results is hidden in the global measure of immediate recall performance.

We examined two other measures, clustering and retention, that are associated with recall in SM in order to understand the nature of the AD deficit in SM. Clustering scores, corrected for chance expected clustering, provide a measure of the extent that semantic organization is used. Both the observed cluster scores and the ratio cluster scores of observed divided by expected performance are presented in Table 1. The AD group showed little or no clustering over the five learning trials, which is consistent with the finding of poor SM recall. Semantic clustering by the NC group, in contrast, progressed over trials, in line with the observed increase in recall performance. The analysis of variance was carried out on the ratio cluster scores to control for different levels of performance between the two groups. There was a significant difference in clustering between the two groups, $F(1, 24) = 107.76, p \leq .001$. These data provide further support indicating that a large part of the
Table 1. Observed and Corrected Semantic Clustering in NC and AD as a Function of Immediate Learning Trials.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Observed Cluster Score</td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>1.92</td>
</tr>
<tr>
<td>AD</td>
<td>.31</td>
</tr>
<tr>
<td>Ratio Cluster Score (Observed Divided By Expected)</td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>1.77</td>
</tr>
<tr>
<td>AD</td>
<td>.43</td>
</tr>
</tbody>
</table>

difference between AD and NC on the overall measure of immediate recall is due to the patients’ impairment in SM.

Delayed Recall
Retention of learning was assessed by comparing free recall after a short (5 min) and long (20 min) filled delay with recall on Trial 5. As indicated in the data in Table 2, the NC group showed minimal forgetting after both delays. In contrast, the AD patients were impaired in their performance after a short and long delay relative to NC. The AD group retained very little of what they had learned, which is consistent with the results reported earlier showing that their recall during learning was primarily from PM.

The analysis of variance of mean recall after a short delay relative to performance on Trial 5 resulted in significant effects at $p \leq .001$ of groups, $F(1, 24) = 64.12$, trials, $F(1, 24) = 110.83$, and the two-way interaction between groups and trials, $F(1, 24) = 20.36$. Similar results were obtained for mean recall after a long delay compared to recall performance on Trial 5; there were statistically significant effects at $p \leq .001$ of groups, $F(1, 24) = 59.36$, trials, $F(1, 24) = 92.83$, and the interaction between groups and trials, $F(1, 24) = 40.56$. Again, these data support the conclusion that the performance of AD patients was based mostly on recall from PM.

Table 2. Mean Words Recalled (SD) on Trial 5 of the CVLT and after a short (5 min) and long (20 min) delay in NC and AD.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Recall Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 5</td>
</tr>
<tr>
<td>NC</td>
<td>11.77 (2.71)</td>
</tr>
<tr>
<td>AD</td>
<td>5.85 (1.82)</td>
</tr>
</tbody>
</table>

DISCUSSION
The results reveal that PM was intact in patients with mild to moderate AD on two different measures: Tulving and Colotla’s (1970) measure which is considered a relatively “pure”

measure of PM, and mean recall from recency based on the serial position curve of the CVLT (Delis et al., 1987). The AD patients’ impairment in overall performance in immediate recall was attributed to a severe deficit in SM. Recall barely improved with practice over five trials.
Not surprisingly, their semantic clustering score during learning was low, and they showed little retention over a 5-min and a 20-min delay.

Our finding of normal PM in AD patients despite severe SM deficits contradicts previous reports of impaired PM in AD patients. As noted earlier, one reason for the discrepancy is that most previous studies based their estimate of PM on tests of short-term recall that are contaminated by SM. These include forward digit span, the Brown-Peterson distractor task, and free recall from the recency part of a list of words. The lower scores that AD patients obtained in the other studies may have been the result of their impaired SM rather than poor PM. Another reason is that even when more appropriate measures of PM were used (Wilson et al., 1983), the severity of dementia either was not properly controlled or at least not specified. In our study, where care was taken to include only patients who were judged to be mildly or moderately demented, PM was comparable to that of normal elderly subjects matched for age and education.

The Tulving and Colotla procedure for estimating PM was used in the present study because it was considered a theoretically more precise measure of PM. An even more precise estimate of PM can be achieved by instructing subjects first to recall the last few words before attempting to recall any of the others. This procedure, however, is different from the "free" recall procedure that is recommended for use on the CVLT. Another potential limitation in using the CVLT is related to it being a categorized list. On free recall, items that are separated during presentation but related semantically may be recalled together, which would result in a lower estimate of PM in normal people with intact semantic processing.

It is unclear whether there is a clinical advantage to using a procedure such as Tulving and Colotla's (1970) to estimate PM. In our study, their procedure yielded similar estimates of PM as the mean recall from recency measure. It should be noted, however, that problems have been reported in the reliability of the recency measure when an alternate form of the CVLT was tested (Delis et al., 1991). Perhaps the Tulving and Colotla measure may be more useful in that regard and can serve as a general tool for assessing patients with other neurological disorders.

An important implication of our findings relates to the clinical need to distinguish between decrements in memory due to normal aging from the abnormal changes caused by dementing illness, such as AD. The present results indicate that PM may be resistant to the deleterious effects of AD at the earlier stages of the disease, as it is in normal aging. Therefore, impairment in PM cannot be used with confidence as a diagnostic indicator of AD at disease onset when clinical diagnosis is critical. Instead, PM may be a marker for impairment at a more advanced stage in the progression of the disease. This may be a fruitful line of inquiry in future studies.

The finding of a deficit in SM is consistent with previous reports of impaired learning in early AD (e.g., Kaszniak et al., 1986). A deficit in SM, coupled with preserved PM, is also observed in amnesic patients who have suffered medial temporal lobe damage with specific involvement of the hippocampus and surrounding cortex (Scoville & Milner, 1957; Squire, 1992). Also, we found that early AD patients, like temporal lobe amnesics (Smith & Milner, 1989; Squire, 1982), are severely impaired in learning and retaining new information when measured by free recall (Freed, Corkin, & Cohen, 1987) even over short, interference-filled delays. The similarity between AD and amnesic patients is not surprising in view of the fact that the medial temporal lobes, which are damaged in amnesia, are the locus of extensive neuropathology in AD (Hyman, Van Hoesen, & Damasio, 1990).

A few clinical implications of our results are worth noting. The finding that PM is normal in early and moderate cases of AD indicates that it is a poor marker for early diagnosis of AD. If PM becomes impaired as the disease progresses, as some evidence suggests (Miller 1971; Morris, 1986), it might serve as a diagnostic indicator of a relatively more advanced dementia. Precise indicators of the various stages of the disease would be useful in addressing clini-
cal issues related to helping patients and their families cope with and manage the cognitive impairments.

SM is a far better indicator of memory loss in AD and, perhaps even more telling, is the ratio of SM to PM in the CVLT. We noted that patients with early AD recalled equivalent or fewer items from SM than PM on the first trial of the CVLT. In many cases, only items from PM were recalled. None of the normal control subjects showed this pattern; their recall from SM was always greater, often by a large factor. Wilson et al. (1983) report a similar pattern in their AD patients. Differential recall from PM and SM may prove to be an important clinical marker to discriminate individuals with AD from people with normal age-related changes in memory and from patients with depression.

REFERENCES


Journal of Psychiatry, 139, (9), 1136-1139.